IMPLICATING VISUALIZATION IN THE RATIONALIZATION OF ALGORITHMS

A Thesis
Presented to
The Academic Faculty

by

Peter Polack

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
College of Computing

Georgia Institute of Technology
May 2017

IMPLICATING VISUALIZATION IN THE RATIONALIZATION OF ALGORITHMS

Approved by:

Dr. Yanni Loukissas, Advisor
School of Literature, Media, and Communication
Georgia Institute of Technology

Dr. Polo Chau, Advisor
College of Computing
Georgia Institute of Technology

Dr. Lauren Klein
School of Literature, Media, and Communication
Georgia Institute of Technology

Dr. Anne Pollock
School of Literature, Media, and Communication
Georgia Institute of Technology

Date Approved: May 2017

ACKNOWLEDGEMENTS

This project's effect on me could not have developed without the presence of my family, friends, and colleagues, who thoughtfully supported this work by caring for my investment in it. I am also humbled and changed by the environment and experiences that I encountered over the course of my master's studies. In particular, I am inexpressibly grateful to my master's advisor Yanni Loukissas, who inspired and enabled me to pursue the kind of work that I have been seeking throughout my creative, academic, and professional careers. This particular project and the work that I have invested into it have been continually motivated by his guidance, which helped for me to identify both my potential and my room for growth. My deep appreciation to my master's advisor Polo Chau, for introducing me to the meaning of academic research, for taking the time to mentor me through the visualization research process, and for epitomizing the qualities of a mentor that carefully attends to his students' work, ideas, and futures. I owe the particular outcome of this project to my committee member Lauren Klein, for providing me with the advice that enabled me to develop potential ideas into meaningful intentions, and for inspiring the importance of this research through her work. I am sincerely grateful for my conversations with my committee member Anne Pollock, who saw in each of my ideas an angle or connection to others that deepend my research interests and the considerations of my work. I am greatly appreciative to have had the consultation of my professor Carl DiSalvo, who has encouraged me to pursue this kind of work with confidence.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In critical algorithm studies [44], the "black box" metaphor characterizes algorithmic systems whose decision-making processes are invisible to their subjects, and yet may affect them. The implication of this critique is that the decision-making processes of algorithms are permissible and acceptable because they are altogether invisible. However, this study contends that algorithms are not only permissible because they are invisible, and are rather *rationalized* according to visibilities and representations that manifest in innumerably diverse ways. Such a project demands an attention to the visuality and representation of algorithms, which would follow from existing work in critical algorithm studies that addresses how designed representations of algorithms influence their perception by users, and in turn shape user perceptions of algorithms [28, 81, 86]. Following from this work, I contend that investigating human-interpretable representations of algorithms is key to studying rationalizations and epistemologies of algorithms, and cannot be discounted from such. In particular, this study investigates 'data visualization' as one instance of these representations, specifically in order to implicate the practice of *representing algorithms interactively* in the rationalization of algorithms. Altogether, the purpose of this approach is to better understand how the services and decisions of algorithms are rationalized, and particularly to contend with the prevailing notion that these rationalizations occur because algorithms are invisible or "black boxed." Therefore, whereas data visualization is the principal concern of this study, it also identifies how algorithms are rationalized by social media platforms, exercise trackers, political discourse, and artworks.

To investigate these claims, I develop a collection of algorithms and interactive data visualizations, entitled *Lifestreams*, which I use to analyze a publicly accessible dataset called *StudentLife* [100, 102]. The *StudentLife* project in particular proposes to use algorithms in order to predict and affect the behaviors of students, which is a paradigm manifest also in data-driven healthcare, work performance evaluation, and criminal justice. To Antoinette Rouvroy and Thomas Berns [82], how these algorithms are perceived by their subjects—their students, patients, employees, or criminals—enables an epistemic rationality that they term "algorithmic governmentality," according to which algorithmic automation acts on behalf of subjects without requiring their active participation and reflection. Accordingly, the issue raised by this work is that algorithmic platforms like *StudentLife* enable an epistemology that escapes human comprehension vis-à-vis human judgment, which can be likened to the problem posed by the "black box." However, in acknowledging that algorithms are able to bypass subjects in this way, I am rather interested in how algorithms are rationalized when subjects perceive and interact with them, for which *Lifestreams* is a case in point. In particular, that data visualizations like *Lifestreams* may rationalize algorithms, and yet require users to perceive and participate in the operations of algorithms, warrants consideration.

Turning to algorithmic governmentality helps at first to identify practices of using and working with algorithms in which humans are in fact marginalized from the decision-making processes of algorithms. In this way, algorithmic governmentality suggests that its epistemic rationality is coextensive with an avoidance of subjects. What is effectively revealed by this argument is that *semiotics* is an essential dimension of this movement whether or not the avoidance of subjects is involved. Plainly, semiotics is the study of how meaning is produced, interpreted, and communicated, whether through symbols, codes, or languages. Accordingly, algorithmic governmentality implies that algorithms and data have specific semiotic properties that influence how they are interpreted by their subjects. Through calling specific attention to this claim, this study examines the premises of algorithmic governmentality in order to address how rationalizations of algorithms, whether or not they entail the removal of subjects from processes of algorithmic decision-making, may involve representations like visualizations that are human-interpretable.

Accordingly, the function of *Lifestreams* in this process is to investigate how the programmatic and semiotic properties of data, algorithms, and visualization inform one another. In particular, *Lifestreams* follows from the proposals of "imaginary media studies" [69] and "critical making" [76], which demonstrate how designing, implementing, and experimenting with media and computing systems enables studying relationships that emerge from their designed arrangement. This enables investigating both the process and product of developing *Lifestreams* as an extended interaction between media, algorithmic, and semiotic elements. Therefore, designing and studying visualization in this way is not to prove how algorithms are universally rationalized by visualizations, nor even to refute algorithmic governmentality, which would require a study of how subjects use, perceive, and experience algorithmic systems. Instead, the object of this study is to investigate how and to what extent the formal properties of algorithms and algorithmic visualizations confer to their rationalization.

1.1 Motivation

The assignment of this study is interdisciplinary in scope and as such relies on multiple related bodies of literature and research. In particular, the study operates at the intersection of critical algorithm studies [44] and visual analytics [95], and is particularly concerned with visual and interactive presentations of algorithms to users. Critical algorithm studies represents a diverse array of research regarding the social implications of developing, using, and interpreting algorithms. From a computer science perspective, the field as a whole departs from a conceptualization of algorithms as discrete, bounded systems and proceeds to interrogate the socio-technical apparatuses of algorithms and algorithmic systems, which not only interact with users at the level of interfaces, but also have broader epistemic and political implications. With respect to visual analytics—which integrates human-computer interaction, visualization, and machine learning research into an interdisciplinary field—critical algorithm studies concerns how representations of algorithms are implicated in how algorithms can be used.

Such a topic has direct implications for research in human-computer interaction and interface design, specifically with respect to emerging disciplines like "ethical user experience design" [77]. Ethical user experience design departs from the directives of traditional human-computer interaction as ensuring system "usability," and opts instead for addressing civil liberties and social concerns when developing computational information systems. For example, whereas traditional human-computer interaction favors a "seamless" design paradigm, according to which users should not be overwhelmed by extraneous information about algorithms that operate in the 'background' of interfaces, ethical user experience design might advocate a "seamful" design that sacrifices interpretative efficiency and clarity for transparency [12]. The concerns of both sides of this discourse are especially relevant to work in visual analytics, which must balance at least usability and transparency concerns in designing algorithmic information systems that allow users to perceive and interact with complex algorithmic processes.

This study contributes to the considerations of critical algorithm studies in a way that might be compared to the directives of ethical user experience design. In particular, it questions the notion that algorithms are problematic when their processes are not represented to users, which is reflected in critical algorithms studies discourse as concerns about algorithms being "black boxed." The black box metaphor characterizes a regime of algorithmic system design in which knowledge is produced by algorithms without human scrutiny or oversight. Accordingly, this knowledge—as well as the algorithms that produce it—are rationalized when they do not confer to human investigation and critique. However, this study contends that rationalizations of algorithms and their decisions do not depend on an absolute flight from human perception and control. Rather, algorithms are rationalized in exceedingly diverse ways, each of which may provide a handle for ensuring the accountability and scrutiny of knowledge produced by algorithms, whether according to critical algorithm studies or ethical user experience design paradigms.

What is at stake is a marginalization of the role of discourse and aesthetics in shaping appraisals of algorithms, their purposes, and their effects. Recent appeals to "algorithmic accountability" [26] and "algorithmic transparency" [18], for example, encourage "reverse engineering" algorithms to extrapolate the logics of black box algorithms in order to expose and audit their quantitative biases. As genuine as this directive sounds, this work is only complicit in representing algorithms as rational systems that are perfected when their causal structures are identified: it articulates the function of algorithms in society as intelligible in quantitative terms, when in fact reverse engineering these processes only simulates their interpretability [48]. For example, Quantitative Input Inference (QII) [18] proposes to achieve 'algorithmic transparency' by calculating algorithmic sensitivity or 'bias' to certain inputs, which is a version of reverse engineering. In doing so, QII rationalizes reverse engineering—as well as reverse engineered algorithms—with an appeal to 'algorithmic transparency' discourse. Accordingly, we might begin to interpret these measures of algorithmic transparency and bias as a kind of IQ or digital phrenology, which are measures that either do not account for the entire picture of algorithms or confer to producing it. In doing so, we can begin to identify how other algorithm discourse and aesthetics, like those provided by academic publications and visualizations, are implicated in rationalizations of algorithms.

To this end, a theory of algorithmic governmentality [82] first identifies ways in which algorithms are rationalized by data "collection, processing, and application" procedures that are not transparent and escape human scrutiny. It suggests in particular that algorithms enable an episteme that diminishes opportunities for the critique of knowledge produced by algorithms, which this study likens to "black box" discourse. In order to evaluate this claim, a semiotic framework is used to identify how these algorithmic procedures may in fact lend themselves to human interpretation, and questions whether algorithms ever recede absolutely from human perception. Accordingly, this study implicates visualizations, interactive games, artworks, and other representations of algorithms in the rationalization of algorithms and knowledge produced by algorithms. Crucially, the theoretical frameworks invoked in this study are not meant to be taken as a matter of fact; instead, they are leveraged to highlight certain dimensions of user interactions with algorithms and their rationalizations.

CHAPTER II

BACKGROUND

2.1 Algorithmic Governmentality

Antoinette Rouvroy and Thomas Berns propose a theory of algorithmic governmentality in order to characterize a kind of rationality inspired by algorithms that anticipate and affect the behaviors of subjects [82]. The notion of governmentality was proposed by Michel Foucault during his project to develop an understanding of the specific techniques and practices that enabled control over the "conduct of human beings" [33]. Whereas Foucault identified how these techniques evolved with respect to certain epistemological and institutional paradigms—e.g., the diagnosis of insanity [32], the development of penal law [34]—governmentality was meant to generalize these techniques and their rationalities as involving an "encounter between techniques of domination over others and techniques of the self" [78]. Accordingly, Foucault carefully differentiates governmentality from the common conceptualization of governance as an articulation of the "state," which Foucault takes as a monolithic and potentially reductive abstraction. Instead, governmentality characterizes the formation, the heterogeneity, and the inherence of techniques of behavior control in enabling the formation of a working society.

Considering Foucault in light of the extensive authority of algorithms in modern day culture, Rouvroy and Berns draw into question the seamless applicability of a Foucauldian tradition of governmentality to algorithmic techniques of behavior control. To accomplish this, Rouvroy and Berns propose algorithmic governmentality in order to identify what techniques of control are uniquely enabled by algorithms [82]. In particular, Rouvroy and Berns propose that what makes algorithmic governmentality unique and differentiable from other forms of governmentality is that it circumvents subjectification, or the processes by which individuals autonomously construct themselves according to their social and cultural environments. Foucault conceptualizes subjectification in order to position "techniques of the self" as techniques of governmentality and behavior control that are not exerted by physical force but rather performed by individual subjects according to normative cultural practices (e.g., marriage), vocabularies (e.g., "you" and "I"), and social ideologies (e.g., party politics). Characteristically, "subjectivation" and "subjectification" involve, respectively, the construction and concomitant self-construction of individual, reflexive subjects according to dominant social codes, a process which Rouvroy and Berns contend that algorithms can supersede without actively involving subject participation.

Therefore, on acount of the automatic reflexivity of algorithms to data, Rouvroy and Berns suggest that algorithmic governmentality advances Foucault's "regimes of truth" to a more immediate "regime of action" [84], according to which algorithms begin to act on behalf of subjects without requiring their active participation. Special to this process is the capacity of algorithms to represent data about individual subjects as statistical "doubles" of them, which are representations of individual subjects in data that are separate from them and yet can be operationalized to make decisions about them. This separation of subject and data enables algorithmic governmentality to deal exclusively with patterns in data that may or may not veridically characterize subjects themselves: "algorithmic governmentality thus focuses not on individuals, on subject, but on relations" between data variables [82]. Accordingly, Rouvroy and Berns suggest that an era of algorithmic governmentality contributes holistically to the "rarefication" of subjectification [82]—an avoidance of subjects and their subjective reflections—which is supported altogether by the classic, techno-materialist conceptualization of computation as powerful because its logic is inherently invisible and therefore obfuscatable [54]. To algorithmic governmentality, whereas statistics enables the development of inferences about subjects that can be applied to them thereafter, algorithms implement an automated, actional, and reflexive statistics that that can operate without human input.

The features of what Rouvroy terms “digital behaviorism” [83] are worth describing here in order to elaborate how a critique of algorithmic governmentality is different from that of statistics. Rouvroy defines digital behaviorism as an episteme “deserted by empirical and deductive, casual logic” and “in favour of computational, pre-emptive, context- and behaviour-sensitive management of risks and opportunities” [83]. To Rouvroy, this epistemology is enabled by algorithms, which can represent unbounded uncertainties in terms of probabilizable risk, operationalize subjects according to these terms, and thereby obfuscate the difference between these uncertainties and their operationalized probabilities. In doing so, the algorithmic techniques of digital behaviorism diminish the possibility of “critique,” which Foucault defines as “a practice that suspends judgment and an opportunity to practice new values,” opting instead for fully automatic risk pre-emption and aversion [82]. The argument leveled here—with its recourse to Foucault and implication of probabilizable risk—finds a comparable precedent in Tiqqun’s “Cybernetic Hypothesis” [97], which addresses the generalization of an algorithmic rationality to contemporary modes of governance and economics. Accordingly, the contribution of digital behaviorism is to identify how the representational and actional capabilities of algorithms contribute to a new regime of decision-making that departs from traditional statistics and confers to a new paradigm of governmentality.

Crucially, the arguments proposed by algorithmic governmentality and digital behaviorism are not about whether algorithms are accurate or verifiable but the extent to which “relying on the apparent operationality of algorithms spares us a series of individual and collective perceptual cognitive, evaluative, conventional, institutional, linguistic efforts or tasks, and, at what price” [82]. Algorithmic governmentality thus concerns the potential and increasing closure of subjectification processes and dissensuses normally provided by scientific or juridicial rigor, instead delegated to the reflexes of algorithmic operations without necessary scrutiny. Such a critique falls also in line with an elaboration of "datalogical turn" [15] as the modern epistemic shift from traditional sociology to computational social sciences, whereby "representation and its reliance on sociological correlation and correlative datasets" is superseded by algorithms that "seek to prehend incomputable data and thereby modulate the emergent forms of sociality in their emergence" [15]. Here, whereas "representation" implies an iterative and subjective critique of evidence, the algorithmic capacity to "prehend" suggests a removal of human deliberation and interpretability for the sake of reflexive immediacy in a "regime of action." Therefore, the function of identifying an algorithmic govermentality or a datalogical turn is not to dismiss algorithms entirely on ethical or nostalgic grounds, but to scrutinize how algorithmic operations enable a suspension of criticality as well as concomitant strategies of behavior control.

2.1.1 Three Stages

Rouvroy and Berns describe three "stages" of algorithmic governmentality that exemplify how algorithmic operations affect the behavior of subjects and are rationalized in doing so: data collection, processing, and application. As opposed to indexing discrete strata of algorithmic governmentality, these stages characterize "blurred" types of algorithmic techniques that, altogether, "are actually all the more effective because they are blurred" [82]. Here, "blurred" means that algorithmic processes across data collection, processing, and application are interactional and entangled, and therefore that they do not lend themselves to decisive deconstruction, taxonomization, or interpretation. Yet, the function of describing algorithmic governmentality according to successive stages is to demonstrate that it is not specific algorithm designs or operations that enable algorithmic governmentality, but rather paradigms of algorithmic use that influence how algorithms are interpreted or obfuscated. Therefore, individually and holistically, the stages can be read to imply a dimension of semiotics, which here concerns the perception and construction of algorithmic meanings by subjects. Considered in this way, algorithmic governmentality becomes less a consequence of algorithmic operations than an effect of how these operations are articulated and perceived. This perspective will be examined in more detail in the following section.

2.1.1.1 Data Collection

Despite popular discourse that treats data as natural and prior to human design [74, 78, 2], data cannot exist without human intervention. Following from Johanna Drucker's proposal to redefine "data," or that which is given, as "capta," or that which is captured [27], data have no intrinsic form and are rather seized and wrought according to technical constraints and designed objective-driven imperatives [61]. Yet, characteristic to algorithmic governmentality is a conceptualization of data as exactly the opposite: an inherent, passive, and a-political aggregation of traces that exist naturally in the world, prior to human intervention. For Rouvroy and Berns, this inherently positivistic viewpoint enables algorithmic governmentality to collect data about subjects without their scrutiny.

Rouvroy and Berns attribute the permissibility of data collection to the fact that initially collected or "raw" data is fundamentally dispersed and prefigurative [82]. Initially "devoid of any prediction about specific end uses," data are "perfectly innocuous [and] can remain anonymous" in the way that their collection bears no immediate consequences for subjects. Moreover "they also appear as absolutely ordinary and scattered," more easily interpreted as atomic traces than explicitly designed or coherent representations. Thus "It follows that we are quite readily give [data] up, for they bear no meaning" insofar as they have yet to be analyzed and operationalized. The a-signification of raw data is therefore key to enabling the permissibility of data collection: “Raw data do not resemble, nor keep even indirect physical bound with any thing in the world, and they are not conventional symbols thereof either,” and they “function as de-territorialized signals, inducing reflex responses in computer systems, rather than signs carrying meanings and requiring interpretation” [83] Here, “de-territorialized” refers to a quality of meaning that is not attached to any discrete structuration of language or social codes, which is opposite to “re-territorialization” or codification [23]. Therefore, algorithmic governmentality suggests that raw data enable a departure from discrete meaning that is complicit in the rarefication of subjectification.

Ultimately, what is enabled by data collection according to algorithmic governmentality is a general obfuscation of the purposes of data, which for their part are collected before they are made meaningful (i.e., re-territorialized). Such a conclusion should be considered with due attention to the temporality and processuality of data analysis [78], which enables data to pass prefiguratively and innocuously into instrumentalization by algorithms ex post facto. To Rouvroy and Berns, this effect culminates in the illusion of data as “traces left and not data shared, though they do not seem to be ‘stolen,’” which suggests a circumvention of subjects and processes of subjectification [82]. With respect to this last point, Rouvroy and Berns are especially adamant: “Together, all these factors eliminate or at least conceal any end goal; they minimize the subject’s involvement, and therefore the consent which can be required for this information sharing, thus removing all forms of intentionality.” Here, the avoidance of “all forms of intentionality” in a regime of algorithmic governmentality presents a more aggressive move than the “rarefication” of subjectification, which might suggest a varied and heterogeneous relationship between specific algorithmic techniques and accordant processes of subjectification.

2.1.1.2 Data Processing

Data processing (also analysis, analytics) tends to begin with an assumption that information inheres in and can be made to emerge from data [74, 78, 2]. Therefore, data processing is a practice of not only experimentally drawing inferences from data but also iteratively shaping these data so that they are commensurable and conducive to analysis. Provided this flexibility, Rouvroy and Berns emphasize that inferences produced by algorithmic data processing can be "uninformed by any pre-existing hypothesis" and yet still operationalizable [82]. Louise Amoore and Volha Piotukh [2] identify the same pattern with respect to "knowledge discovery," which characterizes a set of algorithmic techniques that automatically generate statistical inferences from large swaths of heterogeneous data variables. Amoore and Piotukh describe how knowledge discovery is motivated less by the intrinsic value of relationships between data variables than by statistical confidence and commensurability. They stage this paradigm with a memorable analogy: if big data is the inert substrate that data analytics crystallizes into information, then "little analytics" is the partial lens through which data is rendered valuable, although inauthentic to its totality. Therefore, for both Rouvroy and Berns and Amoore and Piotukh, the truism that "correlation is not causation" is incidentally dismissed by algorithms that can derive value from operationalizing correlations irrespective of their veracity.

Correspondingly, Rouvroy and Berns describe that automation contributes to the permissibility of data analytic inferences, which can be operationalized without human intervention. But what is more is that how these inferences are determined or operationalized can be determined by algorithms as well. Rouvroy and Berns suggest that "The purpose of what is called machine learning is ultimately to directly enable the production of hypotheses based on the data themselves" [82], which implicates the role of self-designing algorithms in further removing human judgment from the jurisdiction of algorithmic operations. Indeed, whereas an algorithm automates data processing, a machine learning algorithm automates how data processing is designed. This feature is significant to Rouvroy and Berns because it suggests a further expulsion of subjects from decision-making in an era of algorithmic governmentality. Rather than enabling an alarmist stance toward the lack of human control in a regime of "artificial intelligence," this focus on machine learning might help to identify exactly what kinds of control are forfeited to algorithmic automations, and to what extent this is justifiable. Altogether—and this is essential—machine learning algorithms do not operate exclusively without human oversight, but they do pardon human involvement at certain points in data processing practice.

Ultimately, automation also entails a practice of generalization that undergirds the operational logic of all algorithms. Although an algorithm might entail a sequence of operations that is only executed one time, the value of algorithms in data processing is that they can execute the same or similar operation for many data consecutively. As an emergent result, data processing produces generalizations—generalization of one operation to many data—which not only enable algorithmic governmentality to operate but incidentally establish these operations as normative and inherent. Whereas algorithms are called "biased" when their generalizations do not satisfy some statistical criteria [51], the generalizations afforded by the inherent biases of algorithms also have a fundamental utility. They can be used to identify normative or anomalous patterns in data, which reveals the relative value of data elements. However, this valuation of data incidentally marginalizes the relative biases of algorithms themselves. Consequently, following from Rouvroy and Berns, "Norms seems to emerge directly from reality itself" [82], even though they are shaped absolutely by the specific biases of algorithms.[1]

[1] For more on this see Ned Rossiter's "determination of relevance" [79], Alexander Galloway's "protocological control" [38], Maurizio Lazzarato's "economy of the possible" [59], Tiziana Terranova's "network dynamics" [94], Geoffrey Bowker's "practical politics" [11], or Michel Serres "parasite" [88],

2.1.1.3 Data Application

The final stage of algorithmic governmentality concerns how data, once collected and processed by algorithms, is applied to subjects. As said, Rouvroy and Berns propose that these data and inferences can be applied to statistical “doubles” instead of to subjects directly, in such a way that “they cannot perceive it, but it is nevertheless applied to them” [82]. The function of a “profile” here is essential, because it is the computational representation of an individual subject that can be altered without a subject’s awareness, and yet has direct implications for the subject’s conduct. Modeled to reflect a subject’s “preferences, intentions and propensities” [51], a profile is especially important for algorithms to anticipate circumstances, whereby information is saved in the present in order to act on it in the future. Such a pattern is the impetus for policy regarding data retention or “oblivion,” which concerns whether and to what extent data about subjects should be permanently kept or eventually expired [9]. If an individual’s past actions are used to represent the nature of their current conduct, then the possibility of self-determination and self-directed deviation from the past is altogether diminished [78]; indeed, such a paradigm entails the rarefication of subjectification.

Importantly, Rouvroy and Berns find the specific circumstances of data application to individual behaviors “less relevant” to a theorization of algorithmic governmentality [82]. This move could be due to the fact that, if algorithmic governmentality depends on the rarefication of subjectification, then a theory of algorithmic governmentality is altogether less applicable to a stage of data application to individuals, which might certainly involve more subjective deliberation than during data collection and processing. To an extent, this explanation is evidenced by the fact that Rouvroy and Berns are keen to emphasize particular scenarios of data application

---

some of which will be returned to below.

that might circumvent subjective reflection entirely, like direct, immediate "action based on the anticipation of individual behaviors [which] could in the future be increasingly limited to an intervention on their environment [and] which is already the case at least during the significant part of life that individuals spend online" [82]. Indeed, this emphasis on the potential of algorithms to modulate environments and interfaces directly strategically circumscribes where data application might be seen to bypass subjectification entirely. Incidentally, this scope sidelines the involvement of representation and subjective reflection in articulating the inferences of algorithms to subjects in a stage of data application.

In addition, Rouvroy and Berns explicitly differentiate between, on the one hand, data application to subjects that might involve their awareness, and on the other, outcomes of data application which need not involve subjectification altogether, like knowledge production. Specifically, they emphasize a difference between "information *at [the] individual level*, on the one hand, which more often than not is observable or perceptible by the individual concerned, and on the other hand, *the knowledge produced through the profiling*" [82]. That is, a distinction is made here between the actional application of algorithmic inferences to subjects (e.g., by physically changing their behavior) and the systematic, usually prefigurative and anticipatory designation of individual subjects to their inferred profiles (e.g., "as a potential fraudster, a consumer, a potential terrorist, a student with high potential, etc." [82]—note the emphasis on potentiality). In doing so, Rouvroy and Berns suggest that data application can happen as a form of knowledge production without actually being applied to subjects. This move warrants focused consideration because it raises questions about how this produced knowledge is ultimately articulated, whether to subjects of algorithmic governmentality or to its operators. If data application is a matter of knowledge production, then how is the knowledge articulated and perceived once produced?

Is subjectification yet involved? These questions are important because they provisionally implicate additional techniques—as well as representations—of algorithmic governmentality that are not implicated by Rouvroy and Berns. Nevertheless, these additional techniques would not necessarily refute the rarefication of subjectification concerned.

Altogether, each stage of algorithmic governmentality exemplifies how algorithmic operations can be used to know about, to make inferences about, and to act upon (or to act on behalf of) subjects without their active involvement. And when the stages are combined these effects are compounded: "their normative functioning is rendered especially powerful and processual by the fact that they mutually reinforce one another (further concealing end uses, further reducing any possibility of intentionality, adapting to our own reality even more, etc.)" [82]. Rouvroy and Berns propose that algorithmic governmentality is thus differentiable from the Foucauldian tradition of governmentality and regimes of truth—Rouvroy contends that algorithmic logic "does not need to go through any ordeal, any investigation, or any exam" in order to be operationalized in practice, which sets it apart from truth established according to "institutions and devices that engaged individuals" [84]. In turn, this "regime of action" obfuscates the purpose of raw data, the difference between real uncertainty and algorithmic probability, and indeed the inherence of algorithmic governmentality in operationalized algorithmic systems, all the while diminishing the possibility of their critique.

2.2 Algorithmic Semiotics

Semiotics, which concerns the study of how meaning is produced, interpreted, and communicated, can be viewed as fundamental dimension of the critique staged by algorithmic governmentality. For instance, we see the semiotics of algorithmic governmentality implicated throughout its three stages: the a-signification of raw data enables their collection (data collection), digital behaviorism and the datalogical turn implicate an epistemological shift away from representation (data processing), and knowledge produced via profiling is not "observable or perceptible by the individual concerned" [82] (data application). Most significantly, the rarefication of subjectification proposed by algorithmic governmentality also implies a rarefication of signification: algorithmic operations elude subjects, contemplation, and critique when they are not directly expressed to them. Considered together, each of these points suggests that algorithmic governmentality implicates a semiotics that privileges the role of a-signification in avoiding subjects [83]. This claim is consistent with the above assessment of data application's scope: algorithmic governmentality is keen to emphasize patterns of behavior control that involve immediate and unmediated actions on subjects, when in fact many of these actions might entail layers of representations, significations, or subjective reflections.

Altogether, it would be valuable to consider the particular mechanisms and representations through which algorithmic operations are signified to subjects, whether through designed explanations of algorithms (e.g., visualizations), or through accidental expressions of algorithmic operations that emerge from their failure (e.g., algorithmic infrastructures that become visible when they break down [90]). Such an approach would not challenge a theory of algorithmic governmentality so much as complicate its dependencies on a-signification—or systems of meaning that elude decisive human interpretation—like those of dispersed raw data, automatic machine learning algorithms, and removed statistical doubles. What is at particular stake here is a marginalization of the role of signification in algorithmic governmentality, which risks viewing algorithms and their influence as entailing an inaccessible and impenetrable regime of top-down behavior control.

In many ways, this problem resembles academic discourse surrounding "black boxes," a metaphor cited frequently in critical algorithm studies to characterize the leverage that algorithms have over their subjects due to the invisibility of their logic [43]. That algorithms can make and operationalize inferences about subjects, all the while without making their logic visible or legible, is the impetus for "opening the black box," "reverse engineering" it [26], or decrying proprietary algorithmic systems [70]. As useful as a black box metaphor is for identifying that algorithmic systems can exercise control over their subjects without their seeing it, this critique fundamentally attributes the power of algorithms to their invisibility and inaccessibility. Incidentally, *this emphasis comes at the expense of considering how algorithms produce knowledge and are rationalized even when they are not absolutely invisible.* In particular, whereas the black box metaphor might be criticized for failing to account for the dynamic complexity of algorithmic systems and their diverse relationships to socio-technical assemblages (e.g., algorithm developers, users, applications) [43], it also neglects the importance of the designed representations and significations that come to stand in for this opaque complexity, which represent it as innocuous, benign, or imbue it with human-interpretable meaning.

For their part, Rouvroy and Berns implicate other features of algorithms besides invisibility that lend to their effects and power, and therefore algorithmic governmentality is not reductive like the black box metaphor. For this reason, the theory's emphasis on avoiding signification is not problematic as much as it is strategically extreme: it implicates the latent potentiality of algorithms to avoid subjects and rarefy subjectification. Indeed, the most absolute or 'late' manifestations of algorithmic governmentality might be expected to be absolutely invisible to its subjects, circumventing subjectification entirely. However, considering how the theory avoids subjects by avoiding signification raises the question of how significations of algorithms are nonetheless implicated in their rationalization.

Addressing this question comprehensively, Ned Rossiter proposes a theory of "logistical media" [79] that identifies the role of algorithmic, digital, and material media in processes that coordinate labor. In particular, Rossiter emphasizes attending to the aesthetics and semiotics of work management interfaces, like visualizations of work performance metrics, that not only direct labor according to algorithms, but also shape how laborers interpret themselves according to these algorithmic interfaces [80] (this might be compared to the control and self-control of governmentality, respectively). When these logistical media are operationalized to determine the placement of people and property according to algorithmic parameters and procedures, they also lend themselves to human observation and reflection. For Rossiter, such percepts may in turn lend themselves to identifying logistical fault lines or chokepoints around which new political subjectivities can be developed: logistical media becomes a plane of political contestation when the knowledge it produces or rationalizes is contested. Logistical media theory thus raises the question: what media render algorithms to human interpretation and rationalization?

Rossiter proceeds in collaboration with Soenke Zehle [107] to implicate the role of algorithmic systems in obfuscating laborers' perceptions of their working conditions. Rossiter and Soenke Zehle identify that, whereas historical threats to laborers such as machinic automation were identifiable, describable, and translatable into organized modes of discourse and resistance, the semiotics of algorithmic information systems confound these more traditional conceptualizations and articulations of labor, preventing their contestation. In particular, Rossiter and Zehle call specific attention to how the simplicity of work interfaces in Uber and Amazon Mechanical Turk contradict the platforms' actually very intricate reformulations of labor. In doing so, the Uber and Amazon Mechanical Turk interfaces effectively depart from a "black box" paradigm of concealing algorithms in order to rationalize them, and turn instead to what might be called a "white box"[2] paradigm: they enlist intuitive representations

[2] In cybernetics and information theory, a white box is model designed to represent an unknown "black box" process—if the white box can be designed such that it exhibits the same functionality as the black box, then the two processes can be considered "isomorphic" and practically the same,

of algorithmic systems that incidentally divert attention from the complexities of algorithmic processes. Although the representations of the white boxes are clarifying and altogether isomorphic to the actual processes of black boxes, they are not any more transparent; accordingly, they may obfuscate the ulterior functionalities of algorithms.

Therefore, whereas algorithmic governmentality is typified by algorithms that bypass signification to avoid subjects, Rossiter and Zehle imply that the total enclosure and autonomy of algorithms is not necessary for an algorithmic governmentality to unfold. At the cusp of this contention is Maurizio Lazzarato's explication [59] of Félix Guattari's "mixed" semiotics, which concerns for Lazzarato, tersely put, how computational information systems produce meanings for subjects. For algorithmic governmentality, the importance of this theory is to chart a movement away from subjectification per governmentality proper, and toward a new paradigm of information systems that produces knowledge and subjects. However, following from logistical media theory, I contend that mixed semiotics is better equipped to understanding that rationalizations of algorithmic operations depend not only on the automatic black boxes of algorithms, but also on their imbrications with clarifying and reflective white boxes. Here, I elaborate mixed semiotics according to human-computer interaction (HCI) paradigms, in order to demonstrate what the theory entails for understanding, studying, and designing information systems. Following this description, I return to the three stages of algorithmic governmentality in light of mixed semiotics, and provide a series of examples in order to illustrate how these theories connect.

2.2.1 Mixed Semiotics

Traditionally, human-computer interaction (HCI) conceptualizes interactions between users and information systems as involving a dialogue or volley between users and

even if their inhering mechanisms are different.

information systems: users provide instructions to information systems that are in turn relayed back to users as "feedback." This model also conceptualizes information systems as *tools*—even in the case of wearable technologies these systems are conceived as *prostheses*. The same notion applies to human-centered computing (HCC), which focuses especially on "usability" and "ergonomics" paradigms in order to design interactions that suit users. For example, the famous human-computer interaction touchstone "The Design of Everyday Things" [68] encourages design that is intuitive, responsive, and does not overwhelm users with extraneous information. Therefore, in the context of human-computer interaction (HCI), mixed semiotics departs from a conceptualization of computer interaction as involving a dialogue between users and information systems, as well as a conceptualization of information systems as *tools*. It accomplishes this in four primary ways which are described as follows, each of which has direct implications for rationalizations of algorithms. Altogether, these dynamics are illustrative and should not be taken as a universal model of meaning, but rather as a "metamodel" [40] that reveals certain relationships and can be set aside once used.

Firstly, because mixed semiotics is a post-structuralist or "critical" semiotics [42], it aims not so much to determine which signifiers produce certain meanings or entail certain referents (like a *structuralist* Sausserian or *pragmatic* Peircean semiotics), but rather to identify how meanings are *socially* constructed. With respect to human-computer interaction, this encourages interpreting information systems not according to the content or meanings that interfaces present to users, but according to how this presented meaning was constructed by multiple actors and algorithmic processes. Accordingly, a concomitant function of mixed semiotics is to identify how constructed meanings engender certain power dynamics. That is, if information systems can produce knowledge, then they can also influence their users according to this knowledge. Lazzarato likens this pattern to an "economy of the possible" [59] which characterizes systems of meaning that define and constrain possible human actions and behaviors, whether according to the knowledge produced by categorical identities (e.g., race, political affiliation) or by information systems.

Secondly, and accordingly, mixed semiotics suggests that interactions between users and information systems cannot be interpreted exclusively in the terms of structuralist semiotics like *linguistics*. That is, although information systems may use language to convey meanings to their users, information systems also use non-linguistic representations (e.g., visualizations of data points) to convey meaning. In order to illustrate this, mixed semiotics invents the hypothetical categories of "signifying semiotics" and "a-signifying semiotics," and likens them to linguistic meanings and "machinic" meanings, respectively. The purpose of doing this is not to suggest that machines can understand meaning (as in artificial intelligence), but to indicate that not all meanings can be deconstructed according to human-interpretable grammars and syntaxes. For example, Patricio Davila [20] likens a-signifying semiotics to *Gestalt* principles in data visualization, which characterize visual relationships among data elements that indicate meanings to users pre-attentively, without recourse to language. Such a non-linguistic and pre-attentive articulation of meaning is one property of a-signifying semiotics.

Thirdly, mixed semiotics suggests that constructed meanings are never stable but always in flux and "mixed" into broader regimes of meaning. It emphasizes the processuality and concoction of signifying and a-signifying semiotics, which are constantly being formed (i.e., territorialized) and de-formed (i.e., de-territorialized) according to systems of meaning and their digression over time. With respect to information systems, for example, the de-territorialized, a-signifying semiotics of raw data may be re-territorialized in the form of human-readable data tables. Therefore, a-signifying semiotics are not so much opposite to signifying semiotics as much as they are atomic constituents or building blocks that ultimately give way to signifying semiotics, or are dissolved out of them [49]. Thus mixed semiotics is supposed to characterize a dynamic processuality instead of a fixed taxonomy of meaning. In the context of mixed semiotics, therefore, "regime" means "a system or planned way of doing things" [46] that can be contested. This suggests, for example, that information systems can centripetally structurate a system of meaning into a meaningful "focal point" [59] that is presented to users, but any number of fragmented meanings can be derived from it. In this way, rationalizations of algorithms can be understood as never totally inherent or stable but in a state of constant contestation and negotiation.

Lastly, mixed semiotics suggests that human interactions with a-signifying semiotics engender a special paradigm of meaning-making called "machinic enslavement".[3] Machinic enslavement can be meaningfully compared to Theodor Adorno's description of radio as an art of enslavement enabled by "atomized listening" [1]: both theories concern a regime of human behavior mediated by "atomized" interactions with media. For mixed semiotics these atomized interactions come in the form of de-territorialized, a-signifying semiotics. To Lazzarato [59], machinic enslavement opposes a "tool-inspired model" of information systems as human *prostheses* in favor of a conceptualization of information systems that generate "modes of enunciation that do not originate in the individuated subject." This effectively implicates the functions of data, algorithms, and interfaces in affecting human behavior according to intricate computational abstractions: "In a machine-centric world, in order to

[3]Because "enslavement" has fervently dystopian connotations, it is worth noting here that Guattari, who coined the term, likened machinic enslavement to the less iniquitous example of driving a car, and he implied the emancipatory potential of a-signifying semiotics: Guattari rejected "anti-modern and anti-machine recapitulations of humanism" that feared a concession of discrete social categories to the fragmented meanings of computation [41]. This principle of Guattari's should be understood in light of his general philosophy and metaphysics, which proposed "molecular" [49], "rhizomatic" [23], and "schizoanalytic" epistemologies that denounced hierarchical regimes of meaning. In contrast, machinic enslavement might also be seen to inform Gilles Deleuze's notion of "dividuals," which famously characterized a regime of behavior control according to de-territorialized or atomized algorithmic inferences that "sieve" individuals [22]. However, especially with respect to Guattari, machinic enslavement should not be read as a problem or a threat, but rather as a paradigm or model of human-computer interaction.

speak, see, smell, and act, we are of a piece with machines and asignifying semiotics." Machinic enslavement thus characterizes how human-computer interactions can rationalize algorithms by rendering their operations as immediate, whereby the meanings of algorithms appear inherent or are interpreted pre-attentively.

2.2.2 Return to Three Stages

Following from this explication of mixed semiotics, Rouvroy and Berns suggest that a-signifying semiotics enable strategies of governance differentiable from subjectification: as opposed to privileging the "representative functions" [82] of language that lend themselves in turn to discourse and reflexive subjects, a-signifying semiotics "synchronize and modulate the pre-individual and pre-verbal elements of subjectivity by causing the affects, perceptions, emotions, etc. to function like component parts" [59], via machinic enslavement.[4] Therefore, mixed semiotics has purchase for Rouvroy and Berns because it emphasizes that subjectification in an era of algorithmic governmentality is changed drastically by the properties of algorithms, which can be considered from this semiotic vantage: given that algorithms can operate without lending themselves to human interpretation or signifying semiotics, a-signifying semiotics are increasingly important in an era of algorithmic governmentality.

On the other hand, Rossiter and Zehle [107] invoke mixed semiotics in order to consider how mixed signifying and a-signifying semiotics interact to produce meanings for subjects, without setting their sights exclusively on machinic enslavement. Indeed, that a-signifying semiotics interact with signifying semiologies in Guattari's model—that they are "mixed" together—suggests that a theory of algorithmic governmentality could benefit from a consideration of how mixed a-signifying and signifying semiotics interact in order to lend algorithms to human interpretation. As

[4] Lazzarato [59] also uses terms like "proto-subjectivities" and "proto-enunciations" in order to illustrate this. These terms illustrate that, respectively, the subjective identities and linguistic enunciations of signifying semiotics are fragmented when they are articulated through a-signifying semiotics.

said, whereas machinic enslavement reflects the capabilities of a-signifying semiotics to affect human judgment pre-attentively, a-signifying semiotics are enmeshed with and “always in the process of being recuperated by a signifying semiology” [49]. Algorithms, for example, elude subjects via a-signification and yet are recuperated in the form of human-readable data tables (e.g., “heart rate data”), the proper names of algorithms (e.g., “T-SNE” [63]), or categorical identities of data subjects (e.g., “good student”). To consider mixed semiotics with respect to algorithmic governmentality, therefore, is to acknowledge the rarefication of subjectification and yet to question whether this trend entails a total departure from signification. For this task, we can return to the three stages of algorithmic governmentality in order to provisionally identify how algorithmic governmentality is enabled not only by significations in human-interpretable media but also by interactions between signifying and a-signifying semiotics.

2.2.2.1 Data Collection

Data collection entails a fundamentally black boxed process that operates in the ‘background’ of computational platforms like web browsers, mobile applications, and credit card transactions. Accordingly, algorithmic governmentality emphasizes that data collection enables an avoidance of subjects, especially when data is collected without an individual’s awareness or active contemplation. However, it is also valuable to consider the role of representation and signification in rationalizing these collections altogether. That is, although the collection of personal data tends to float largely under our radar, as it were, this process and its effects may also make themselves known and interpretable in ways that have recourse to intuitive, signifying semiotics. Furthermore, whereas the implication of data collection in algorithmic governmentality is that data can be repurposed to any end after it is already collected, we might consider how data collection is rationalized in other ways; that is, as opposed to concealing from subjects how data will be used. Such rationalizations of data collection are not limited to obfuscatory "Terms of Service Agreements" (which are for their part hardly human-interpretable), and rather especially include scenarios of data collection in which the data concerned is produced or created by actively participating subjects.

Rouvroy and Berns address that user interactions with social media "are full of *signifying* semiotics" to the extent that "many people have become obsessed with producing subjectivity" [82]. The implication here—and usually the implication of Guattarian signifying semiotics in general—is that signifying semiotics lend themselves to immersive, visceral, and reified meanings that may altogether be inauthentic to some ulterior functionality (or power dynamics). That is, for social media, the human-interpretable and subject-oriented signifying semiotics of being a user with a certain number of 'followers' or 'likes' give way to systems of meaning and concomitant behaviors that reinforce the values of these meanings; simultaneously, these signifying terms mean nothing to algorithms that collect data about them insofar as they deal exclusively with patterns among this data and their a-signifying semiotics. Therefore, Rouvroy and Berns's invocation of social media incidentally suggests how signifying semiotics can enable data collection in algorithmic governmentality to pass innocuously, by concealing the backgrounded functionality of collecting data about interactions with these signifying systems of meaning. This (white box) conceptualization of data collection should be set alongside the prior (black box) one, according to which data collection is permissible because data is dispersed, prefigurative, and without a definitive purpose. Indeed, both signifying and a-signifying dimensions of data collection may comprise the same information systems.

In surveillance studies, Julie Cohen [16] characterizes this pattern of data collection that requires participation from subjects as evidencing a "participatory turn" in surveillance, according to which innocuous tasks of crowd-sourcing mask their ulterior agendas of data collection. Cohen notes in particular the media ecology of Nike+ for representing data collection attractively, specifically by implementing "gamification" in wearable technologies to overshadow the importance of data collected for targeted marketing. Cohen also calls attention to Foursquare as a valorized "map for nothing" [31], which gives its users the opportunity to add anything about their lives to a shareable geography, which consequently collects this data for investors to access and use. In both cases, signifying semiotics are implicated in a process that Cohen pithily characterizes as "playing and being played," by which subjects deal exclusively in entertaining, intuitive, and innocuous systems of meaning that lend themselves to ulterior functions of data collection.

Whereas Rouvroy and Berns note how algorithmic governmentality can be rationalized "under cover of 'personalizing' information" [82], Cohen's Foursquare example further implicates how this "cover" can leverage signifying semiotics. Particularly, as opposed to representing data collection directly, Foursquare represents data collection according to what Cohen notes is a *diegetic* gameplay experience. Accordingly, Cohen likens the diegetic and non-diegetic in games to the subjects and operators of gamified surveillance environments, respectively. That is, whereas the subjects of gamification act within the seamless, immersive, and consistent constraints of a gameplay narrative, operators acknowledge that the non-diegetic game environment enables conditions of action that are bounded within a controlled system. It is such that the participatory turn characterizes designing a narrative for the subjects of data collection, which lets data collection pass in a way similar to that proposed by algorithmic governmentality. However, the diegetic gameplay experience further implicates the role of signifying semiotics in rationalizing data collection: data is not only dispersed and prefigurative—it is contemporaneously territorialized in the form of Foursquare locations that users create and interact with.

Furthermore, the participatory turn entails a broader rationalization of data collection according to what Cohen terms the "surveillance–innovation complex": an ideological symbiosis of agendas for data collection and imperatives for technological innovation, in which "participation and commodification are entwined as a matter of political economy." As opposed to simply enclosing data collection processes in a black box, the surveillance–innovation complex engenders a new political subjectivity that privileges the value of "crowd-sourcing" and "open data" in enabling a prosperous economy and society. Therefore, in addition to Nike+ and Foursquare, Cohen describes how information policy legislation underscores a rhetoric of "information processing as innovation," whereby data innovation, openness, and autonomy is put in opposition to government regulation, which is said to suppress innovation. That these trends might be read as technoscientific variations of neoliberalism suggests that subjectivity indeed has a significant part to play in rationalizing algorithmic approaches to data collection and processing.

The surveillance–innovation complex can be compared to algorithmic governmentality in that both critiques enable a shift away from deconstructing the mixed semiotics of interfaces—as in machinic enslavement and gameplay diegesis—and toward interpreting broader, epistemological rationalizations enabled by mixed semiotic regimes. This is to view mixed semiotics as, on the one hand, 'mixed' at the level of interfaces, algorithms, visualizations, and games, and on the other, consistent in enabling a broader algorithmic rationality that is not instantiated in code and yet manifests in discursive, diagrammatic, and indeed legislative articulations of algorithms. Although this distinction risks dichotomizing and reifying how a mixed semiotic framework should be applied, it is useful because it prevents reducing mixed semiotics to a map of the material architectures of algorithmic systems. In other words, signifying and a-signifying semiotics should not be exclusively understood as, respectively, the visible 'front-ends' and algorithmic 'back-ends' of computational ‘stacks’; instead, computational semiotics also interact with broader socio-technical assemblages that are manifest in other mixed semiotics. This enables a departure from a conceptualization of information systems as conveying meaning to users in an isolated or ‘closed’ system.

Accordingly, considering data collection in light of mixed semiotics enables a consideration of how broader social customs and cultural practices—for example, in the form of “open data” legislation—enable data collection to pass permissibly through the diegetic mixed semiotics of algorithmic interfaces. And such an insight is not opposed to a theory of algorithmic governmentality: that data are intrinsically dispersed, prefigurative, and a-signifying may strengthen the capability of signifying semiotics to relate this lack of territorialized meaning to permissible, enticing, or addictive data collection narratives. At the same time, and as suggested by a surveillance–innovation complex, we might further witness the emergence of new norms regarding which kinds of these articulations are permissible themselves. Therefore, what is also important to attend to here is how regimes of data collection are normalized over time or otherwise disputed. For example, although data collection processes may not ask users explicitly for permission to collect their data [98], this behavior is also being flagged and contested as problematic [39]. Therefore, how this interaction paradigm is rationalized is not only a function of diegetic mixed semiotics in a closed system, but also one of broader social norms and mixed semiotics.

2.2.2.2 Data Processing

Whereas data collection meets subjects at their smartphones or wearable devices, we have insofar characterized data processing as the absolutely a-signifying, automatic, and invisible logic of algorithms. And according to Rouvroy and Berns’s critique of machine learning, even the design of this logic can be determined by black boxed algorithms that escape subjective reflection and critique. Such an argument has its merits, specifically in acknowledging how algorithmic logic may be even more permissible when it is opaque [60]. However, a focus on mixed semiotics encourages us again to consider how data processing lends itself to various systems of meaning that are not so easily characterized as ‘black’ or ‘white’: rationalizations of algorithms depend not only on whether their operations are visible to users, but also, per mixed semiotics, how these visibilities are constructed.

As established by the field of “visual analytics” [95], which concerns the interdisciplinary design and study of machine learning, human-computer interaction, and data visualization, the need to render algorithms and their operations intuitively and accessibly is heightened by the increasing complexity of their functions and inferences. In particular, even though machine learning algorithms leverage information in data to determine how to best operationalize it, this process is itself highly supervised and controlled by users, at least according to human-interpretable statistics about how these algorithms satisfy some statistical performance criteria (classifier precision or recall are fundamental examples). Accordingly, visual analytics is one particularly direct means of rendering data processing for human interpretation and use. Although visual analytics is examined extensively later in this study, it is worth noting here that the explicit object of visual analytics is not to rationalize algorithms, but to lend them more accessibly to human interpretation. Therefore, we will later examine how this directive may nonetheless rationalize algorithmic operations.

On the other hand, Christian Sandvig [86] implicates the function of visual, sonical, and even theatrical representations of algorithms in rationalizing data processing operations according to enticing abstractions. Sandvig calls attention to metaphors of data processing—an animated assembly line that explains Google Search (Figure 1), a vintage industrial machine that represents Facebook EdgeRank (Figure 2)—for shaping appraisals of algorithmic systems according to connotations of comprehensiveness and objectivity. The reader will note that these metaphors rationalize post-industrial ecologies in the image of mundane Fordist paradigms, which is remarkably consistent with Rossiter and Zehle's critique of Uber and Amazon Mechanical Turk. With respect to mixed semiotics, a caricature of an algorithmic system as an industrial machine leverages signifying semiotics to rationalize the intricate logics, dynamics, and design considerations of algorithms. Similarly, Sandvig implies that the definition of "algorithm" is indeterminate, and more a function of its collective representations in diagrams, advertisements, and metaphors. Indeed this would have it that the very definition of "algorithm" is one of mixed semiotics. Therefore, when Sandvig calls for a "counter-visuality" that would represent algorithms in different, non-visual terms, he appeals to the contestation of this mixed semiotic regime.

If rationalizations of data collection can occur both at the level of interfaces (e.g., diegetic games) and at the level of political discourse (e.g., the surveillance–innovation complex), then rationalizations of data processing are not restricted to representing algorithms through interfaces either. Rossiter and Zehle [107] explain how Walmart rationalizes its algorithmic operationalization of labor according to "social physics," which frames data analytics of labor performance as an objective science. Social physics [71] is a mathematical abstraction of interactions among crowds that proposes the development of infrastructure and information systems according to specific measures of crowd dynamics. In the way that it is applied by Walmart, social physics is exemplary of a mixed semiotic regime that structurates meaning according to abstractions, and is used in turn to rationalize algorithmic data processing. Like the surveillance–innovation complex, and very much unlike games or visualizations, social physics does not explain how algorithms work so much as it licenses a rationality of using algorithms. Following from mixed semiotics, we can interpret this regime of meaning as something that is sometimes territorialized into information systems, advertisements, or labor policy, or is otherwise contested.

Figure 1: The Google Search algorithm represented as an animated assembly line (designed by [45], excerpt by [86]).

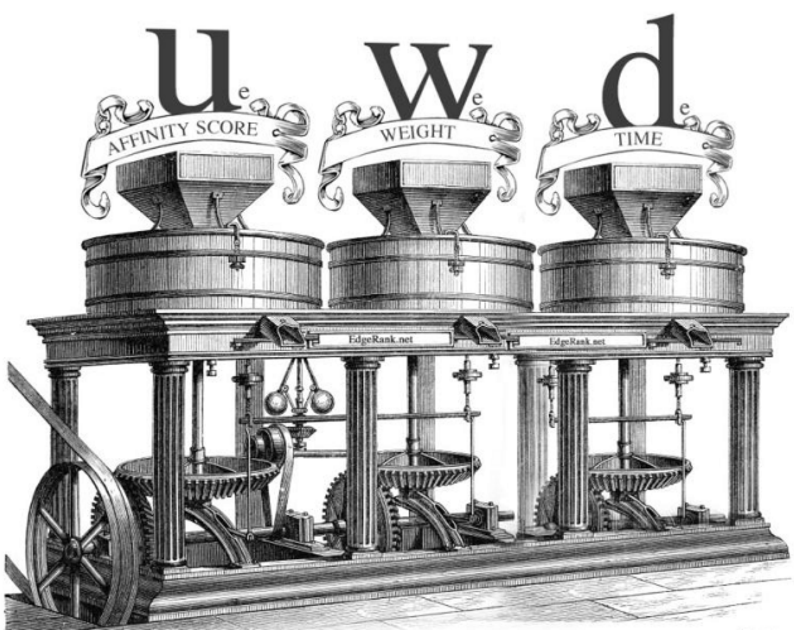

Figure 2: The Facebook EdgeRank algorithm represented as an industrial machine (designed by [104], excerpt by [86]).

2.2.2.3 Data Application

Whether data application can occur without subjectification or signification is questionable. Even in the case that data application produces knowledge about subjects without their awareness, the fact that this knowledge must eventually be interpreted, used, or applied suggests that data application can never remain absolutely removed from subjective interpretations. Granted, this may be one reason that Rouvroy and Berns propose a "rarefication of subjectification" and not an absolute elimination of subjectification. Indeed, such a complete avoidance of subjects from data application is theoretically possible, but it would require a kind of 'total' machinic enslavement, like one presaged by cyberneticians or technological determinists. Nonetheless, for the time being we might attend to the ways in which data application is rationalized by systems of meaning that are perceived by their subjects.

Initiating work in this direction already, Ganaele Langlois [57] applies mixed semiotics to deconstructing the Amazon.com web interface, which reveals how the a-signifying semiotics of algorithms mix with the signifying semiotics of interfaces in order to construct coherent meanings for users. Amazon's representation of books is particularly interesting to Langlois, as they demonstrate case in point how the signifying semiotics of book cover images, page contents, and pricing information integrate with the a-signifying semiotics of how these books are defined programmatically in data. In addition, Langlois calls attention to Amazon's book recommendation system, which algorithmically extrapolates similarities between user profiles in order to suggest catalogs of books to users based on their interests. Whereas this system evidences aspects of algorithmic governmentality, it also entails a mixed semiotics in which the fragmented data of user profiles is territorialized into the cohesive signifying semiotics of books that are "Recommended for you, Peter." Indeed, that algorithms compute book recommendations is trivial, but that these algorithmic processes are meaningful according to subjective names, identities, and interests is a consideration enabled by mixed semiotics. Accordingly, Langlois's application of mixed semiotics to Amazon book recommendation exemplifies a paradigm of rationalizing algorithms during a stage of 'data application' that does not avoid subjects, and in fact addresses them directly.

However, recalling the relationship between diegetic mixed semiotics—like at the level of interfaces—and broader mixed semiotic regimes—like at the level of political identity, legislation, and discourse—it becomes clear that Langlois's Amazon example evidences a tendency toward the former. That is, although the study decisively deconstructs the Amazon recommendation system according to the diegetic mixed semiotics of Amazon's web interfaces, less conspicuous in her study is how the mixed semiotics of Amazon book recommendations are rationalized by systems of meaning beyond the scope of diegesis, like interface aesthetics or interaction norms. An example of this might include the fact that recommendation systems are typically evaluated according to their faults, not only by their developers, but also increasingly by journalists, and victims of inappropriate or otherwise uncanny recommendations [3]. This broader awareness or 'literacy' of the implications of recommendation systems can be productively viewed as a mixed semiotic regime that compromises the stability of the system of meaning manifest in Amazon book recommendation. Such is a function of the mixed semiotics framework: to acknowledge that these systems of meaning, no matter their instantiation in code, are constantly being upset and re-negotiated.

Another regime of meaning that influences the rationalization of algorithmic recommendation and information systems is one of aesthetics. To reiterate, "regime" here means "a system or planned way of doing things" [46] that can be contested. Therefore, to invoke aesthetics with respect to mixed semiotics follows most closely to the work of Jacques Rancière [75], who proposes in a theory of "aisthesis" that art should be characterized not by distinct phases of artistic genres and periods, but by evolving canon of aesthetic and representational invention which is being continually disrupted. Therefore, aesthetic regimes are being continually reinvented and contested in a "distribution of the sensible." How might this relate to rationalizations of algorithms? Rossiter and Zehle [81] invoke "aisthesis" explicitly in describing an "aesthetics of algorithmic experience," which characterizes a similar paradigm of aesthetic regimes, but according to the human experience of living among algorithmic representations and valuations of life. Examples of this already discussed include social physics, as well as metaphors of algorithms as machines, both of which are part of a collective regime of meaning—discursive and aesthetic—that rationalizes the application of algorithms to subjects.

In this way, an aesthetics of algorithmic experience might be read altogether as an *aesthetics of algorithmic governmentality*: how algorithmic governmentality is rationalized according the prevailing aesthetics of algorithm representations, applications, and effects. Furthermore, in this way, even mundane representations of 'data processing' and 'data collection' are implicated as aestheticizations of 'data application' and algorithmic governmentality. Like Cohen, Rossiter and Zehle [81] note the media ecology of Nike+ for its partial representation of algorithmic purposes and effects. The two examples they cite are useful for illustrating how data application is rationalized when it is aestheticized: whereas the in-store visualizations of the Nike+ running app depict data collected from wearable devices in an organic 3D interface (Figure 3), the data-driven statues of the Nike China Logistics Center represent energy savings and logistical efficiency as animated LED displays (Figure 4). Therefore, whereas algorithmic governmentality traces a movement of rationalizing algorithms according to the removal of subjects from algorithmic decision-making process, these examples suggest an inclination to, nonetheless, rationalize these processes through highly visual and subject-oriented aesthetic displays. For example, we might add to this list "beautiful" visualizations of cybersecurity operations [65] (Figure 5) or algorithmic artworks as

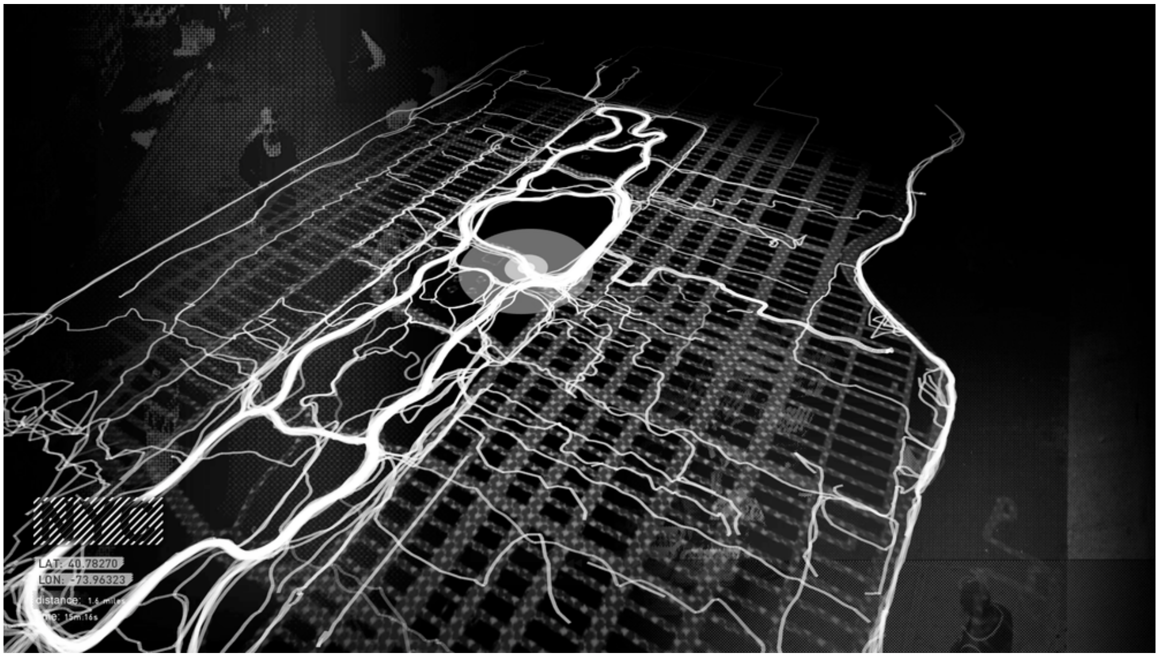

Figure 3: The Nike+ *City Runs* visualization (designed by [73]) depicts data collected from wearable devices as an venous grid of geographic activity. The question raised by Rossiter and Zehle is: how does this representation confer to a certain understanding of data collection algorithms?

popularized by Google's neural network "hallucination" projects [64], which bear no interpretable meanings to public audiences and yet rationalize algorithmic operations according to highly stylized a-signifying semiotics. The purpose of these aestheticizations are to present algorithmic applications to the public, but not necessarily to render these applications accessibly.

Through invoking Rancière in a similar manner to Rossiter and Zehle, Alexander Galloway [36] contends that visualizations are, unlike art, constrained to a particular aesthetic regime. Galloway likens visualizations to a "banality of representation"—like the aesthetic compliment of Hannah Arendt's banality of evil [4]—which implicates computational "modes of production" for determining the representational capabilities of visualizations. For example, beyond data points and graphs, there are some non-quantitative aspects of data—e.g., how subjects experienced the process of data collection—that visualizations are simply unable to represent. Therefore,

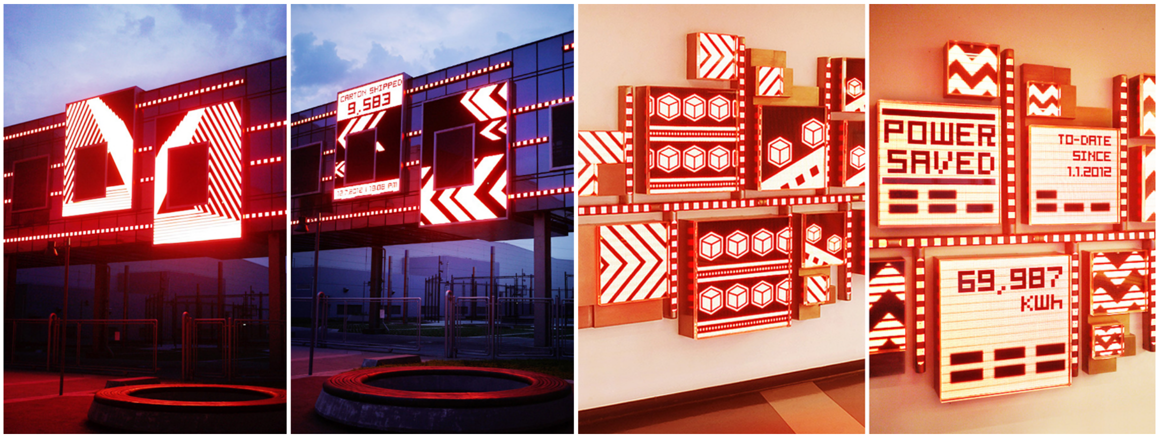

Figure 4: LED displays (designed by [24], photographed by [25]) at the Nike China Logistics Center aestheticize energy savings and logistical efficiency metrics.

Figure 5: The *Daedalus* cybersecurity visualization [65]. Although spheres appear to represent networks and red pop-out glyphs cyberattacks, the visualization is unintelligible to a public audience without insider knowledge, and is rather functional as an aestheticization of cybersecurity operations.

whereas Galloway sides with Rancière, Rossiter, and Zehle to acknowledge how visualizations entail a "distribution of the sensible" [75], Galloway pushes back on the opportunism that a new aesthetic regime can happen on behalf of artistic ingenuity alone. In this way, Galloway implicates the effects of computational materiality on representational and aesthetic possibilities. Alongside an aesthetics of algorithmic experience, this together depicts the total rationalization of algorithms as an evolving "aisthesis" that is yet constrained by the representational capabilities of visualizations. Therefore, if an algorithmic "regime of action" [82] exists, it is continually rationalized by algorithm aesthetics and concurrently constrained by their modes of production.

Altogether, somewhat unlike for data processing, representations of data application are less concerned with the logic, procedurality, and functionality of algorithms than with the implications of algorithms for human experience. To be sure, data collection, processing, and application are artificial categories imposed here for the sake of illustration—they are "blurred"—but segmenting them in this way helps to move from algorithmic games to algorithmic visualizations to algorithm aesthetics, the latter of which characteristically avoid representing the materiality of algorithms [81]—e.g., whether logistical efficiency at the Nike China Logistics Center is achieved by optimization algorithms, labor policies, both, or neither—in order to articulate the vague epistemic rationality of an "algorithmic culture" [92]. And yet, this departure from materiality does not yield a total black box; instead, what substitutes a depiction of algorithmic *logic* (i.e., data processing) is a vibrant and persuasive show of algorithmic *effects* (i.e., data application). This implicates an aesthetics of algorithmic governmentality that cannot be reduced to the diegetic mixed semiotics of information systems.

2.3 Discussion

Following from a theory of algorithmic governmentality, we identified how properties of algorithms and their data enable the rationalization of algorithmic inferences, especially in a way that diminishes opportunities for critique and subjective reflection. We described data collection, processing, and application as "blurred" algorithmic techniques that evidence aspects of algorithmic governmentality, which consequently implied the function of semiotics in rationalizations of algorithms. Then, according to the mixed semiotics framework, we characterized rationalizations of algorithms as systems of meaning that reify algorithmic inferences, whether according to human-interpretable representations of algorithms (i.e., signifying semiotics) or real-time human-computer interactions with algorithmic systems (i.e., a-signifying semiotics). We identified gamification, visualization, and book recommendation as instantiating *diegetic* mixed semiotics in this way, and we described how appeals to innovation, metaphorization, and aestheticization engender broader mixed semiotic *regimes*.

To be clear, operating algorithms in real time through visual interfaces does not refute the possibility of algorithmic governmentality nor its rarefication of subjectification. In fact, the participation of analysts in operationalizing algorithms does not include the actual subjects of algorithmic governmentality, but rather its operators. And yet, this caveat raises the question about who the subjects of algorithmic governmentality include—presumably the subjects of algorithmic governmentality are the individuals about whom data is collected and applied—but are algorithm operators and analysts not a kind of subject as well? Altogether, as is made most evident by the "quantified self" [93] and "personal genomics" [53] movements, this distinction between algorithmic subjects and operators is beginning to blur—very much like the stages of algorithmic governmentality—as data subjects are given tools to collect and analyze their own data. Therefore, data processing is not necessarily under the exclusive jurisdiction of algorithms nor professional analysts so much as increasingly diverse kinds of users that interact with and have access to algorithmic systems in different ways. For this reason, algorithmic governmentality is altogether useful for questioning whether there is such a thing is a definitive *subject* of algorithms, since the distinction between behavior control (e.g., healthcare interventions) and self-control (e.g., quantified self) is blurred and being blurred by algorithmic information systems [78]. In turn, mixed semiotics suggests that these control and self-control dynamics implicate systems of meaning that are distributed across subjects, algorithmic processes, and interfaces—and they are never strictly invisible nor hierarchical.

In addition, it is noteworthy that the theories of algorithm representation and rationalization surveyed here call for a "counter-aesthetics" [36] or "counter-visuality" [86] to existing representations of algorithms. These appeals to aesthetic and visual intervention can be read as an argument about the function of signifying and mixed semiotics in rationalizing or disputing the knowledge produced by algorithms, which exceeds and yet supports the scope of algorithmic rationality. Rouvroy and Berns [82], for example, are interested in "emancipation" from algorithmic governmentality, which they define according to a subjective dissonance with "seemingly harmonious social tools." In other words, the subject-avoiding mechanisms of algorithmic governmentality can be contested whenever they break down [90] or conflict with the interests of subjects. Following from the notion of counter-aesthetics, we identify that this dissonance paradigm also involves discursive, aesthetic, and political engagements with algorithms.

For critical algorithm studies, this discourse it useful because it invites critical studies of algorithms that are altogether less reactive to the failures of algorithms, and more conducive to exploring their possible implications. It would also invite abandoning the black box metaphor in favor of a more nuanced conceptualization of algorithms that are rationalized by mixed semiotic or aesthetic regimes of meaning. Rationalizations of algorithms do not depend on whether algorithms are represented or not, or whether they are represented accurately or not, but on how heterogeneous "mixed" systems of meaning coalesce to produce intuitive meanings about algorithms to subjects. Indeed, an appeal to counter-aesthetics in light of mixed semiotics, more so than algorithmic governmentality and the surveillance–innovation complex, helps to identify that discursive and aesthetic articulations of algorithms are important: they entail changes in how algorithms are conceptualized.

To conclude, Sandvig's [86] and Galloway's [36] particular arguments question whether designed representations (of algorithms) can at all be unproblematic, especially with respect to visualizations. This is a valuable assessment because whereas algorithmic governmentality problematizes a-signifying semiotics, these critiques of visualization problematize signifying semiotics as well. Therefore, the problem of a-signifying semiotics proposed by algorithmic governmentality will not simply be resolved by signifying semiotics, and indeed these mixed semiotics must be conceived as mutually interactional, if not mutually enforcing. 'Opening the black box' will only yield a white box that rationalizes algorithms in a different way.

CHAPTER III

LIFESTREAMS

Considered together, algorithmic governmentality and mixed semiotics help to identify that, although the subjects of algorithms can be marginalized from their decision-making processes, the processes, effects, and motivations of algorithms are articulated to subjects with recourse to diverse mixed systems of meaning. Accordingly, we might turn to visualization as one instance of these systems, and study how algorithms are rationalized by their semiotics. Importantly, such an inquiry would not conclusively determine whether algorithmic governmentality exists or whether it is coextensive with a rarefication of subjectification; instead, it would address how visualization is implicated in rationalizations of algorithms. Therefore, this inquiry departs from the specific critique staged by algorithmic governmentality, and considers more so its illustration of an algorithmic rationality enhanced by mixed semiotics. The object here is to identify how visualizations and rationalizations of algorithms are mutually implicated by theories of algorithmic governmentality and mixed semiotics.

In this section I describe the design, development, and analysis of *Lifestreams*, a collection of interactive data visualizations that I developed in order to study and demonstrate the relationships between visualizations and rationalizations of algorithms, specifically with respect to theories of algorithmic governmentality and mixed semiotics. Following from mixed semiotics, this study of how visualizations (or any systems of meaning) rationalize algorithms should not be reduced to a conclusive determination or taxonomization of how specific semiological forms operate to produce meaning. Rather, visualization is interpreted as producing a system of meaning that is not strictly linguistic, can be interpreted by users in other ways than by language, and can be contested. This enables a special attention toward the visualization design process, the algorithm development process, and visualization aesthetics, the latter of which is commonly removed from visualization research. This altogether pushes back against paradigms in human-computer interaction (HCI) that might study how visualizations represent or rationalize algorithms; it instead foregrounds the processual, temporal, and dynamic character of meanings co-constructed among algorithms, visualizations, their users, and their developers.

The purpose of *Lifestreams* is to examine interactive visualization in the rationalization of algorithms, using the three stages of algorithmic governmentality identified by Rouvroy and Berns [82] (data collection, processing, application) as its point of departure. Beyond this, *Lifestreams* is an experiment in expressing the premises of algorithmic governmentality, mixed semiotics, and rationalizations of algorithms to users through an interactive system. That is, whereas *Lifestreams* is used to examine the ideas explored in this study, it also attempts to demonstrate these theoretical relationships more comprehensibly. Altogether, the success or failure of *Lifestreams* depends not on its ability to prove that mixed semiotics or visualizations are always involved or complicit in algorithmic governmentality, but rather to effectively reveal how algorithmic governmentality and mixed semiotics can highlight the ways in which algorithms are rationalized by visualization. Lastly, although *Lifestreams* is not meant to challenge algorithmic governmentality or its rarefication of subjectification, its implications with respect to these ideas will be elaborated in the Discussion.

3.1 Overview

Lifestreams (Figure 6) is a collection of algorithms and interactive data visualizations that can be used to analyze the *StudentLife* dataset [100, 102]. The *StudentLife* dataset was designed, implemented, and procured by a team of computer scientists at Dartmouth College in order to "shine a light on student life" [102] in a way conducive to data-driven analysis. The *StudentLife* dataset contains various data about 48 students' physiologies, academic performance, and opinions collected over a 10-week term, primarily via a smartphone application installed on the students' phones. The *StudentLife* dataset is publicly accessible and downloadable online.

Lifestreams is designed in particular for use by data analysts who want to better understand the Dartmouth student body, to determine which student behaviors contribute to good academic performance, and to identify anomalous behaviors that could put students at risk. Accordingly, *Lifestreams* is equipped with four visualization components (Figure 6) that suit different kinds of data analysis, and are interoperable. The scatterplot uses dimensionality reduction algorithms (T-SNE [63] and PCA [96]) to visualize the *StudentLife* students according to their overall similarities, which allows groups of students to be selected and analyzed. The data table lists all data variables imported from *StudentLife* into *Lifestreams*, allows these data variables to be manually toggled on and off, and lists the average values for the students being analyzed. The location graph visualizes student movements between Dartmouth campus buildings in a directed graph, which is derived from *StudentLife* wi-fi data, and can be customized according to a series of simple ontologies: group by building, group by building type, group by activity type. And the timeline visualizes student temporal data over the course of a 10-week term. All components are linked such that any selection in one is reflected in the others, and the scatterplot and location graph animate in order to emphasize how changes between selections relate to the data.

A typical scenario of using *Lifestreams* involves selecting certain data variables of interest in order to group students by similarity according to those variables. These groups are reflected in the scatterplot, which updates automatically. Groups of students can then be selected from the scatterplot, which highlights their corresponding data throughout the other visualization components (Figure 7). Data characteristics

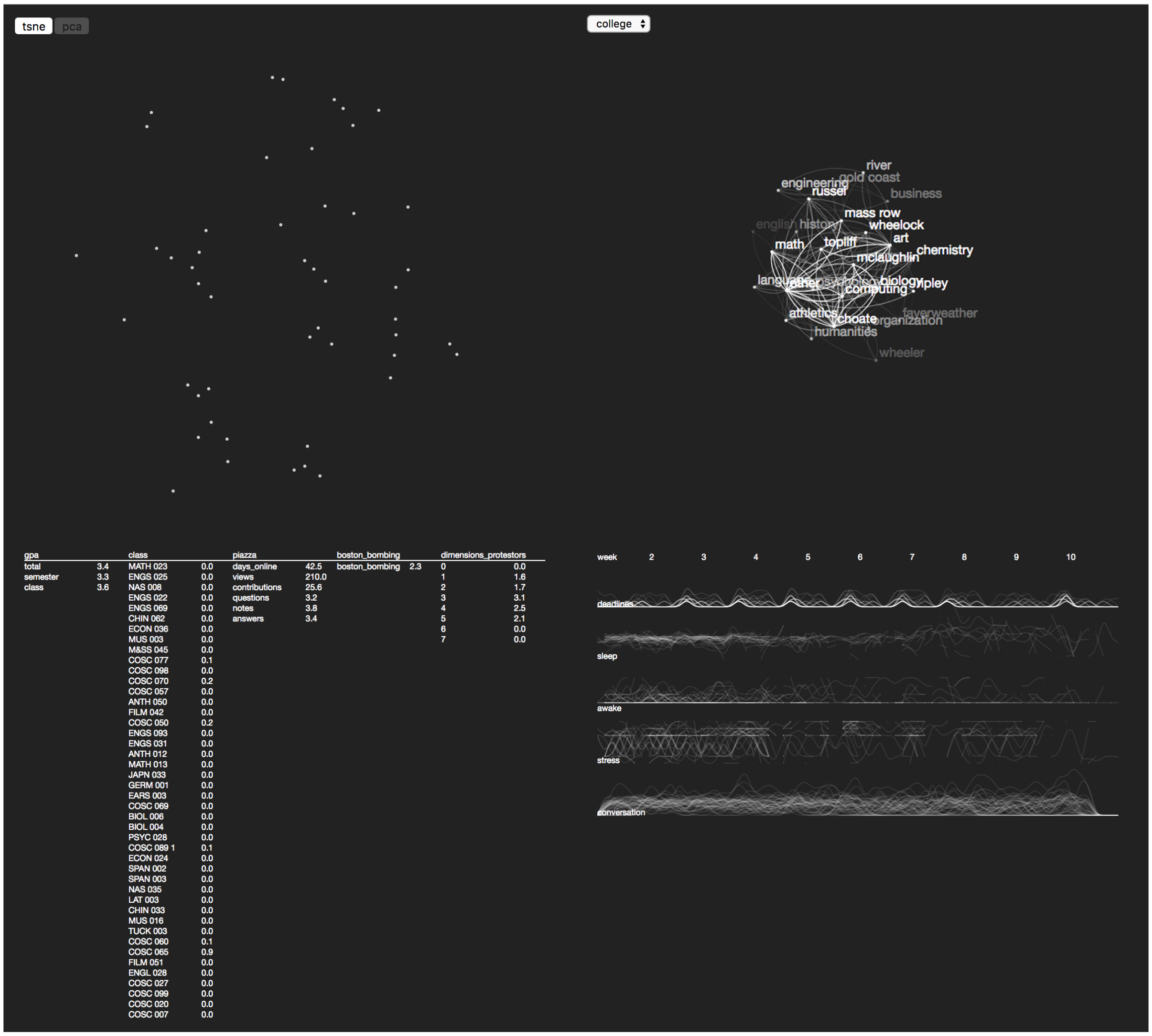

Figure 6: *Lifestreams* without any data selected. The scatterplot (top left) arranges students from the *StudentLife* dataset according to the similarity of their data variables, which can be selected from the data table (bottom left). The location graph (top right, enlarged to show detail) depicts the geographic movements of all students according to their connections to wi-fi hot spots, which are collated here as different Dartmouth academic departments and geographic regions. The timeline (bottom right) depicts student temporal data (deadline counts, sleep hours, ability to stay awake in class, stress levels, and conversation counts). Additional data variables from the *StudentLife* dataset can be imported into *Lifestreams*.

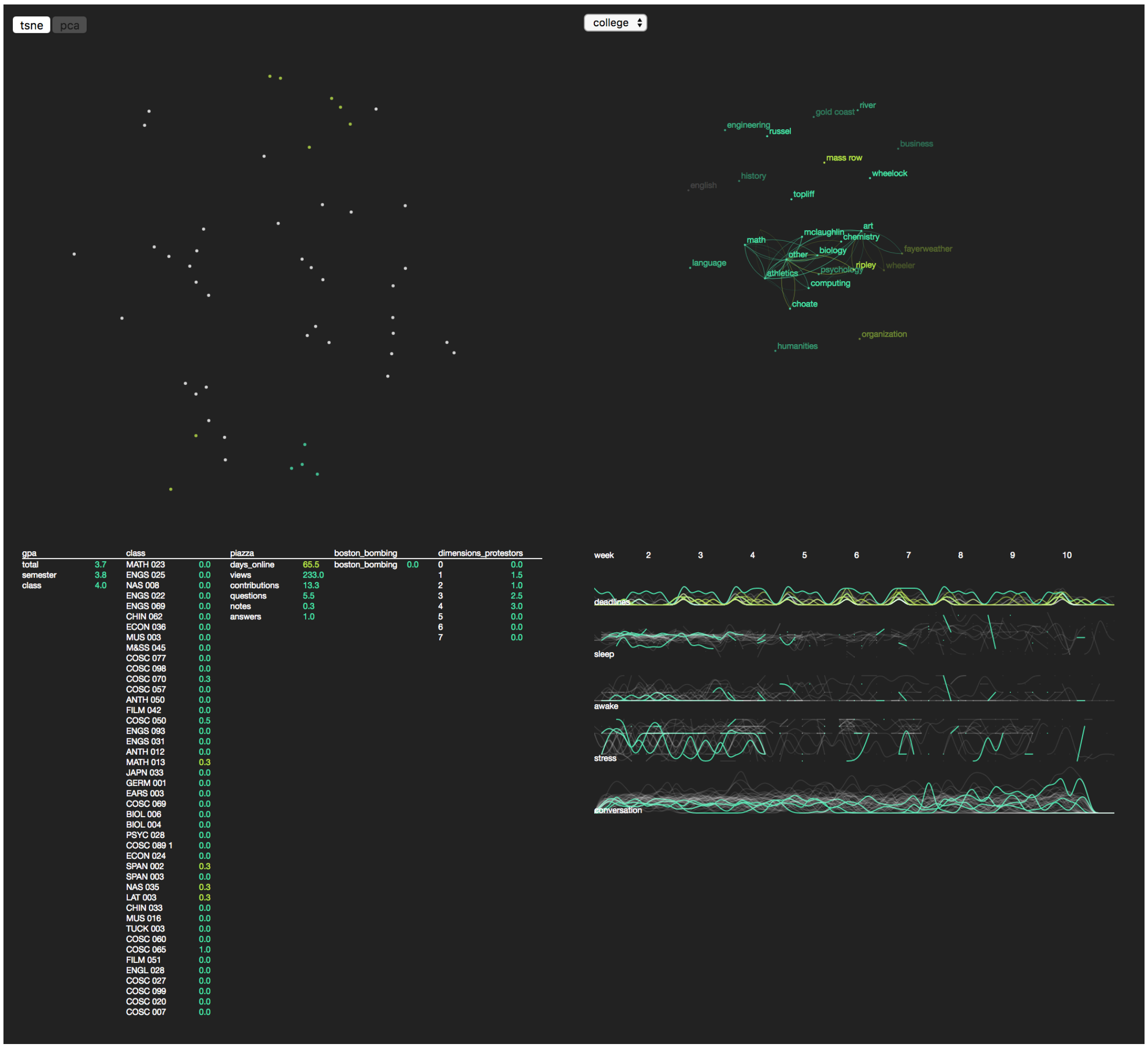

Figure 7: *Lifestreams* with four student points selected in the scatterplot. In doing so, the data table updates to display quantitative information about the selected students. Locations visited by the selected students are highlighted green in the location graph. Data variables and locations that are statistically indicative of the selected students are highlighted yellow, as are other student points that reflect these similarities.

that are statistically indicative of the selection (e.g., one standard deviation from the mean) are made additionally prominent in order to allow for inferring broader relationships between student behaviors. For example, if the selected students have a low cumulative GPA compared to the remainder of the population, then the cumulative GPA value will be highlighted redder in the data table; if the selection frequents Dartmouth's Sudikoff Laboratory relatively more than the remainder of the population, then Sudikoff will be highlighted greener in the location graph. In addition, other students and data that share similar properties to these are highlighted accordingly; for example, other students that have low cumulative GPAs and frequent Sudikoff are highlighted redder and greener. This reveals higher-level statistical relationships throughout the population in a visual fashion that can be successively explored.

Lifestreams was designed and developed according to a data ontology that was created to make the *StudentLife* data more conducive to data analysis. This ontology organizes the *StudentLife* datasets into a hierarchy depending on their applicability to certain kinds of analysis, and is in turn reflected in *Lifestreams*'s metadata scheme, which enables the system to visualize data types in the interface components that suit their analysis. Accordingly, the individual visualization components in *Lifestreams* reflect the structure of this data ontology.

3.2 Precedents

The decision to produce a visualization in order to research it is informed by a diverse body of research in design theory, media studies, critical algorithm studies, and computational pedagogy. Making visualizations and algorithmic systems like *Lifestreams* affords critical algorithm researchers the opportunity to trace information systems not only according to the content that they represent to users, but also according to their processual development. Therefore, this analysis concerns not only visualization as a product, but also visualization as a process and a practice that reveals certain dimensions of how algorithms are rationalized. Meanwhile, and following from appeals to designing 'counter-aesthetics' [36, 86], *Lifestreams* operates to study how making visualizations can contribute to an original understanding of algorithmic rationalizations. This is a highly reflexive task that follows from a series of 'research-through-design' [108] and 'research-creation' [13] precedents, which broadly entail the use of a design practice to investigate or produce research theory.

Lifestreams can be positioned at the intersection of imaginary media studies [69], critical making [76], and procedural literacy [19]. Imaginary media studies aims to develop fictional alternatives to existing media platforms in order to study them, and critical making emphasizes a practice of playing with this space of alternatives through activities of hands-on construction. In both cases, the experience of producing, witnessing, or discussing designs takes precedence over establishing their immediate functional utility. Similarly, procedural literacy encourages the design of critical and pedagogical platforms that enable experimentation with procedures, such as the logical processes of algorithms. Although procedural literacy's definition of procedure extends beyond the scope of algorithmic logic, it nonetheless tends to promote instantiating procedural rules in interactive, modifiable, and therefore typically computational environments. *Lifestreams* is both a speculative and procedural design project because it envisions an imaginary regime of using algorithms, and yet it enlists real data and visualization tools in order to do so. And it approaches the directives of procedural literacy by enabling users to experiment with the statistical and logical procedures of algorithms.

Some pedagogical platforms that approach this intersection between speculative and critical design, critical making, and procedural literacy already. Matt Ratto and Jean-François Blanchette's "It's a Series of Tubes" [87] workshop invites participants to design a network communications protocol for interacting with a car track set, which teaches about the computational design of packet flows and their politics. Carl DiSalvo and Jonathan Lukens's "Neighborhood Networks" [62] program adopts a critical and speculative design approach to developing technological literacy. Tom Jenkins's and Ian Bogost's "Tiny Tinkering Platform" [52] promotes accessible experimentation with devices in an 'Internet of Things.' Elements from each of these projects are evident in RYBN's "Antidatamining" [85], which offers a pluggable platform for experimenting with the mechanisms of algorithmic trading. Each project, developed for procedural literacy, critical making, or speculative design, uses design—hands-on or imagined—to identify, articulate, and make accessible the implications and effects of algorithmic operations. Such a focus, especially for

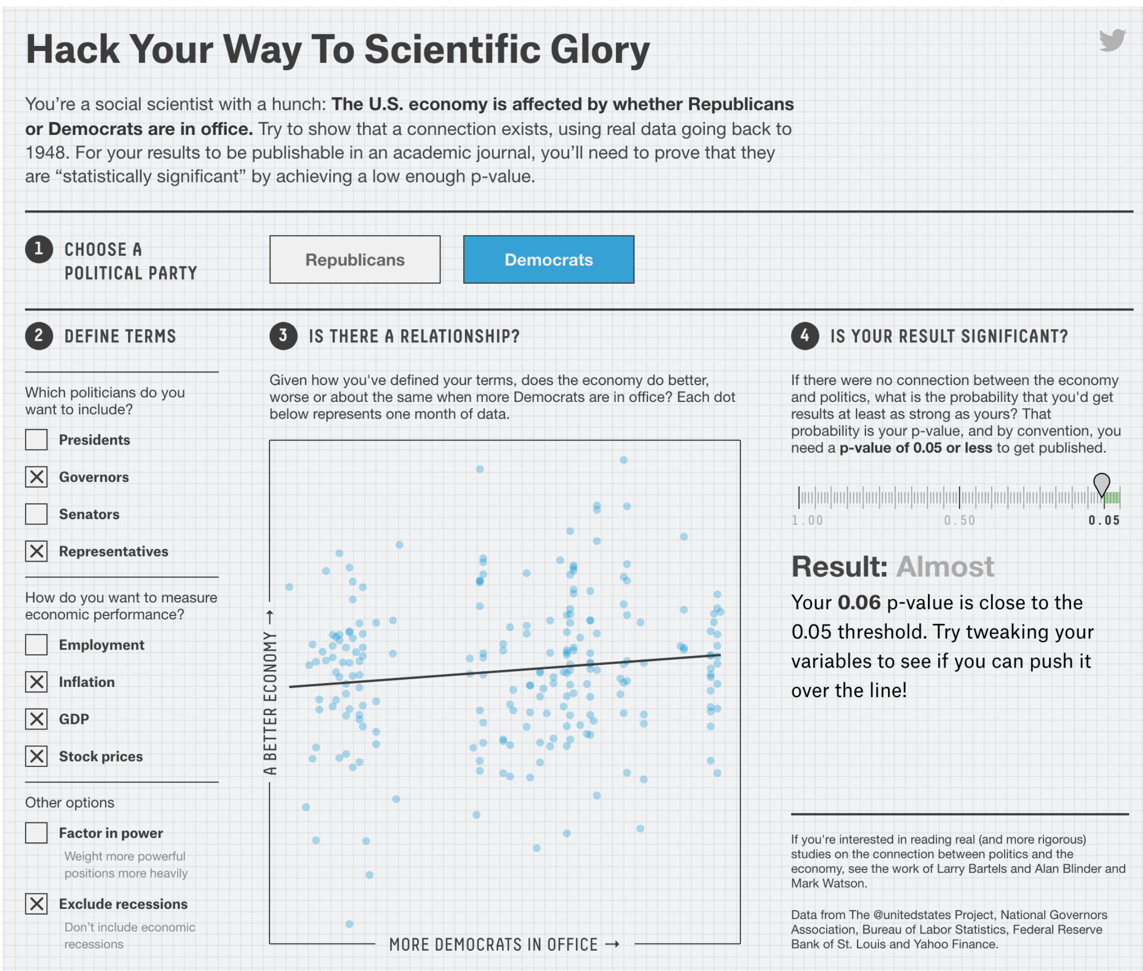

Figure 8: Interactive visualization of "p-hacking" [5], wherein users experiment with variables in order 'cherry pick' statistically significant correlations.

“Antidatamining” in particular, is also to attend to a dynamical or “feral” model of algorithms which reveals patterns that emerge from interactions with and between algorithms, refusing to reduce algorithms to strictly deterministic engineering problems [35].

More specialized solutions to developing procedural literacy include “educational testbeds” [55] in computer science pedagogies that teach algorithm design through experimentation with programmable agents in virtual environments, as well as “explorable explanations” [99] that have made algorithms visually interactive in order to teach how they work (also see Google TensorFlow’s “Neural Network Playground” [89]). Although the majority of these platforms lack an emphasis on criticality, open-ended design, and the political implications of procedures—indeed these considerations are less trivial to implement into code—some outstanding instances exist. FiveThirtyEight’s interactive explanation of “p-hacking” [5] (Figure 8) demonstrates how data variables can be cherry picked to yield statistically significant, “publishable” correlations. “Serious games” can incorporate politically sensitive material into the content or procedures of video games [10].

Finally, some of Ned Rossiter’s research explores how field studies about logistical media, work performance metrics, and the algorithms that guide them can be rendered as interactive games and speculative interfaces [79]. A noteworthy example of this approach is *Logistical Worlds* [17], which accompanies a documentation about labor management infrastructures with visualizations and games that simulate the scenarios concerned, rendering them accessible to research and reflection. In this way, Rossiter’s work with Soenke Zehle extends the notion of “procedural literacy” to a call for a “parametric politics” [107], which encourages experimenting with algorithmic parameters, procedures, and their political implications. A parametric politics draws inspiration from anonymity and piracy discourse in order to encourage (re)purposing algorithms to political applications, especially in order to reveal the political and epistemic implications of algorithms [21]. Following from Johanna Drucker's call for "aesthetic provocations" in designing computational systems [29], aesthetics should be understood as a dimension of this form of experimentation. Such a consideration is supported altogether by the theories of mixed semiotics and algorithm aesthetics addressed in this study.

3.3 Method

Lifestreams follows from "imaginary media studies" [69] and "critical making" [76] in order to investigate how designing, implementing, and experimenting with algorithms enables studying relationships that emerge from their designed arrangement. It draws particular inspiration from research projects such as *Antidatamining* [85] and *Logistical Worlds* [17], which leverage a practice of both implementing and exhibiting visualizations in order interrogate properties of data, algorithms, and their relationships to visual representation. This particular approach enables investigating both the process and product of developing *Lifestreams* as an extended interaction between media, algorithmic, and semiotic elements. Accordingly, *Lifestreams* was developed according to three steps that were designed in order to study the process and product of visualization development. First, I obtained and studied an existing dataset to examine its relationship to the 'data collection' stage proposed by Rouvroy and Berns; second, I developed algorithmic and interactive visualizations that implemented elements of the 'data application' stage; and third, I noted how the development and form of these visualizations rationalized algorithms, especially in light of theories of algorithmic governmentality and mixed semiotics. The implications and considerations made over the course of designing these three steps, prior to implementing them, are elaborated as follows.

3.3.1 Dataset

The point of departure for *Lifestreams* was to implicate interactive visualization in the rationalization of algorithms, which could be framed according to the three stages of algorithmic governmentality identified by Rouvroy and Berns: data collection, processing, and application. Following from the precedents discussed in the section above, I used a publicly accessible dataset to implement a visualization that would speculatively contextualize how 'real' data might be purposed and used. Although 'synthetic' data could be generated that could still be operationalized by algorithms and visualized, the value of 'real' data is to relate these algorithmic operations and visualizations to actual algorithmic subjects and 'data collection' procedures. In particular, the chosen *StudentLife* dataset [102] was produced according to a variety of needs that developed out of project goals, institutional collaborations, and protracted engagements with study participants. Therefore, the *StudentLife* data expresses a rich diversity of real constraints that manifest as idiosyncrasies that might not be evident in 'synthetic' data—including missing data and inappropriate participant responses. These artifacts not only shape *Lifestreams*'s design, but also enable this study to consider a stage of 'data collection,' which occurs prior to the *StudentLife* dataset and yet is manifest in its presentation, structure, and content.

The *StudentLife* dataset was also selected for evidencing aspects of algorithmic governmentality. Whereas the *StudentLife* project involves a 'data collection' apparatus that requires direct participation from its subjects in order to function, *StudentLife* also involves 'data processing' outcomes in which subject profiles are used to produce knowledge about their behaviors according to statistical relations between data variables. Altogether the *StudentLife* project involves an "evaluation of mental health and academic performance" [100] that is correlated to a variety of other heterogeneous data variables. *StudentLife* never 'applies' this data to its subjects, but altogether rationalizes the implementation and use of 'data application' algorithms, which might "provide tips" [100] to students according to inferences about their academic performance. The *StudentLife* project produced a dataset, a data collection apparatus, a body of data processing work, and data application implications that all might be unpacked according to algorithmic governmentality and mixed semiotics.

Although the *StudentLife* dataset was published with a documentation of its data, documenting the data manually presented an opportunity for me to interpret the data content and structure without assistance—as is often required in other scenarios of data analysis—and also to compare findings from my own 'data archaeology' with the official documentation. This process resulted in a data guide that I referred to throughout developing *Lifestreams*, particularly in order to organize its development and design. Furthermore, when I experimented with simple statistical correlations among data variables of interest, I realized the need for a more advanced data ontology, which entailed designing a data structure that was more conducive to data analysis. Altogether, the development of the data guide and ontology informed the actual design of *Lifestreams*—what algorithms it uses, how its visualizations are organized—which suggests the importance of attending to the process of visualization development inasmuch as its outcome.

Altogether, that *Lifestreams* implements a 'real' dataset means that it can be actually applied to analyzing the *StudentLife* data. Therefore, although *Lifestreams* is a purely speculative product that intends to highlight certain aspects of algorithmic rationalizations, it is also a working tool that can be used in practice. Additionally, in this way, *Lifestreams* rationalizes the *StudentLife* dataset and its operationalization by demonstrating that the dataset can be meaningfully used. Whereas the *StudentLife* dataset is already rationalized by its online presentation and academic accreditations, *Lifestreams* joins this ecology of *StudentLife* applications and makes its own claim for how the data might be employed.

3.3.2 Design

Although existing visualizations that evidence rationalizations of algorithms could be studied in practice, the value of developing a visualization from scratch (that is, from raw data) is to reveal how rationalizations of algorithms are implemented or otherwise emerge from the constraints of data analysis. Granted, because developing an algorithmic system from scratch entails a degree of authorial or creative flexibility that comes at the cost of empirical validity, developing and studying *Lifestreams* does not conclusively determine whether algorithmic governmentality or its rarefication of subjectification exists, or whether visualizations need be involved in this process. Instead, *Lifestreams* identifies to what extent interactive visualization may be involved in rationalizations of algorithms, whether in accordance with a theory of algorithmic governmentality or not. In addition, *Lifestreams* retains a significant degree of empirical legitimacy because it is implemented within the constraints of computation and 'real' data. Therefore, as a kind of speculative design bounded by the constraints of 'real' data and computational capabilities, *Lifestreams* is an experiment in realizing a realistic a probable space of algorithmic practice.

If *StudentLife* represents mostly a stage of data collection (as a caveat, data processing is involved in all data collection, and that the *StudentLife* is publicly accessible beckons its application in some manner), then *Lifestreams* would represent mostly a stage of data processing. As discussed, data processing can be rationalized by visual analytic systems that operationalize algorithms, by visual or metaphorical representations of algorithms, or perhaps by pitch black boxes. As this study is particularly interested in the implications of the former, *Lifestreams* is designed for a hypothetical scenario in which a data analyst is determining and operationalizing algorithmic inferences about the *StudentLife* data. To reiterate, whether this analyst should be considered a subject of algorithmic governmentality is indeterminate, but for the purposes of studying rationalizations of algorithms, the general premise of algorithmic governmentality applies. Altogether, *Lifestreams* could certainly be used by its own data subjects (i.e., the *StudentLife* study participants) or made publicly available in a way that would aestheticize or rationalize the directives or applications of the *StudentLife* project. Therefore, that *Lifestreams* is designed for data analysts should not invalidate its implications for data subjects, and these kinds of users may very well be one in the same.

Data application exists in *Lifestreams* only to the extent described by Rouvroy and Berns, whereby predictive inferences are made about subjects without their knowledge. For example, when *Lifestreams* is used to identify "high performers" (i.e., good students), it can be read as exhibiting a kind of data application according to the 'production of knowledge.' Beyond this, and with the exception of how *Lifestreams* could be applied to aestheticize the *StudentLife* project, other kinds of data application—such as actually affecting subject behavior according to algorithmic inferences—are generally excluded from *Lifestreams*'s design. This is due in part to the fact that a speculative implementation of this kind of data application to the subjects of the 'real' *StudentLife* data would be grossly unrealistic and contrived. Although *Lifestreams* does not intend to be realistic, it is supposed to be authentic to how the *StudentLife* data might be foreseeably and reasonably operationalized. Therefore, including this form of data application in *Lifestreams* would require additional considerations about how algorithms would be applied to subjects in the physical world, and how subject behaviors would be affected in real time, which would exceed this study's scope.

An important design focus in developing *Lifestreams* was to enable its users to compare human and algorithmic decision-making, to reveal compromises between them, and to understand their relationships in light of mixed semiotics. This focus prompted the incorporation of a series of algorithms and visualization techniques that effectively related rationalizations of algorithms to the premises of mixed semiotics and algorithmic governmentality. These considerations, as well as how they were informed by the actual constraints of the *StudentLife* data, will be discussed.

3.3.3 Development

As said, developing an algorithmic visualization from scratch presented an opportunity to implicate the development process in rationalizations of algorithms. This demanded keeping a journal throughout this process. Throughout the process of developing the data guide, designing the data ontology, pre-processing the data, and implementing *Lifestreams*, I noted the relationships between algorithmic governmentality, mixed semiotics, and rationalizations of algorithms. However, to reiterate, although this study involves applying a mixed semiotic framework to the codes and structures of data, algorithms, and visualizations, it does not intend to determine how these codes and structures are universally implicated in rationalizations of algorithms. For this, case studies or empirical examinations of existing algorithmic systems might be better suited (as in Langlois [57]). Instead, the function of research-through-design [108] and research-creation [13] in this study is to examine the processuality and procedurality of how visualization is implicated in rationalizations of algorithms, which a mixed semiotics framework can help to reveal.

3.4 Findings

The notes taken over the course of developing *Lifestreams* are elaborated as follows, according to three parts that reflect the development process: Dataset, Design, and Development. Dataset concerns my exploration of the *StudentLife* dataset, its documentation, and its public presentation, in order to determine how *Lifestreams* should be designed. Design describes how I developed a plan for operationalizing this data, which influenced the final design of the *Lifestreams* interface. Development documents my implementation of this plan, as well as the challenges and opportunities raised. Each of these parts includes some notes about how the development process was motivated by algorithmic governmentality, mixed semiotics, and rationalizations of algorithms. However, it is in the following Discussion section that the theoretical implications of these notes are cohesively synthesized and discussed, especially with respect to using the *Lifestreams* interface.

3.4.1 Dataset

Consistent with algorithmic governmentality, the *StudentLife* data was dispersed and prefigurative in a way that avoided particular considerations about its ultimate applications. Indeed, although the *StudentLife* dataset as a collection was not dispersed and prefigurative in that it was structured, organized, and presented as a coherent package, its comprising datasets were often unrelated to one another and were not labeled with any definitive purposes. In other words, the ultimate applications of each individual dataset within the *StudentLife* dataset were undefined, and therefore their implications for students (the data subjects) were indeterminate. As opposed to intending to reveal something in particular about students, the data was collected under the nebulous directive of "studying the health and performance of students" according to "the impact of stress, mood, workload, sociability, sleep and mental health on academic performance" [100], and was later operationalized in various ways according to correlations between data variables (e.g., inferring GPA [103], inferring stress [50]). Some *StudentLife* directives, however, such as being "interested in how students' mood changes during the day" [102], may have influenced the collection of certain data. Altogether, the discursive presentation of the *StudentLife* dataset privileged statistical relations between data variables in data processing over an adherence to hypotheses established in advance.

The *StudentLife* dataset contains a number of individual datasets that each represent a particular data variable[1]: e.g., student stress, sleep, location. The purposes

[1] "Dataset" and "data variable" are used somewhat interchangeably here, although where "dataset" is used to refer to a collection of some data, "data variable" refers to the information

of some *StudentLife* datasets were more clear than others. For example, whereas the `grades.csv` dataset entailed direct implications for understanding student academic performance, it was more ambiguous how the class seating position dataset might be applied.[2] Furthermore, the format and content of the datasets influenced to what extent their operationalization had implications for subjects. For example, whereas the `wifi` dataset contained the same information as the `wifi_location` dataset, the latter represented the alphanumeric identifiers of wi-fi 'hot spots' (i.e., the MAC addresses of access points) as the proper names of Dartmouth buildings. In light of mixed semiotics, this example in particular reveals a contrast between the a-signifying semiotics of `wifi` data and the signifying semiotics of `wifi_location` data, the latter of which can be more trivially related to student behaviors. Like grades data, `wifi_location` data consists of signifying semiotics that have more direct implications for interpreting students according to their data. Futhermore, that certain datasets are more intelligible and applicable in this way might incentivize their use: the *StudentLife* developers deliberately converted the `wifi` dataset into the `wifi_location` dataset in order to render it more intelligibly and accessibly [102]; accordingly, I only used the latter dataset in my analysis.

The meaning and format of each data variable is documented on the *StudentLife* website, and additional information about how these data were collected and derived is available in academic publications posted there. Whereas the titles of the datasets (e.g., `conversation`) already confer to understanding their contents and potential applications, the data documentation on the *StudentLife* website elaborates what

about subjects that a given dataset provides. A dataset (or database, data table) contains one or more data variables (e.g., stress, sleep, location) and one or more data instances (e.g., individual subjects, students). Algorithms, for example, enlist multiple data variables to make inferences about data instances—information about which is provided in data sets.

[2] Notwithstanding the fact that some students manipulated the seating position data collection system to insert website addresses into the data.

these data are, how they were collected, and sometimes how they were algorithmically derived. Given that the datasets are documented to varying levels of detail, the *StudentLife* data documentation can be read as an argument that articulates the overall functions and purposes of the *StudentLife* data collection system and dataset. For example, certain datasets were not described in the data documentation: `app_usage`, `calendar`, `call_log`, `dinning` [sic], and `sms`, which might suggest that these data were collected and yet deemed relatively irrelevant to *StudentLife* goals, or altogether uninsightful or unfit for analysis. In addition, with the exception of `dinning`, these undocumented data variables were related to routine smartphone storage processes, which might account for their omission. Accordingly, the *StudentLife* data documentation might be seen to reflect both the project's goals and the constraints of its system architecture. With respect to mixed semiotics, this relationship could be productively viewed as the way in which the a-signifying constraints of algorithms and algorithmic systems confer to certain signifying semiotics, whether in the form of dataset titles or documentations; in turn, when mixed at the level of the *Lifestreams* interface, these specific relationships become less clear.

Conversely, there were data variables that were documented on the *StudentLife* website but omitted from the *StudentLife* dataset. For example, student Facebook data was removed from the *StudentLife* dataset with the note, "not sure we can release this until we are convinced it is fully anonymized" [100]. Anonymization concerns the removal of personally-identifiable information (PII) [56] from data, which includes information such as personal names, geographic locations, or in the case of Facebook, browsing history or friendships. Interpreting anonymization according to mixed semiotics suggests that data which cannot be cleaved from signifying semiotics is simply not rationalizable under certain circumstances of use, which is determined in part by data collection and application norms. Therefore, removing PII from data, or removing PII data from datasets, entails a removal of signifying semiotics from data. For example, another example of data removed due to anonymity concerns in the *StudentLife* dataset is the “opportunistic face logging” data [101], which captured photographs of student’s faces when they interacted with *StudentLife* questionnaires. Whereas algorithmic applications of this data were unsuccessful [101], the signifying semiotics of students’ faces prevent them from being publicly accessible and analyzable. That the *StudentLife* dataset, like most all data, is stripped of signifying semiotics in this way, should be interpreted in light of the critique developed by algorithmic governmentality: information systems may altogether privilege a-signifying semiotics that are removed from subject-oriented categories.

Ecological momentary assessments (EMAs) [91] are a significant part of the *StudentLife* dataset for collecting data about student opinions, moods, and activities, usually in the form of text-based questionnaires. The *StudentLife* project administered these EMAs eight times a day on average, and was principally concerned with ensuring student “compliance,” which is a measure of voluntary participation in EMA and other data collection procedures [102]. For example, compliance diminished over the course of the semester because students “discarded repetitive EMAs as the novelty wore off,” and so compliance was incentivized by sending email reminders to students and awarding merchandise to top data “collectors” [102]. Therefore, whereas some data collection procedures are generally invisible and occur in the background of students’ smartphones, EMA data collection involves highly visible algorithmic apparatuses that incentivize subject participation. In addition, signifying semiotics in the form of nomenclature like “compliance” and “collectors” ascribe subjective identities to data subjects, in order to incentivize their participation and perhaps to de-incentivize their non-compliance. To note, EMAs also prompt a high degree of subjective reflection from data subjects, which could make this kind of data an exception to algorithmic governmentality.

Documentation about EMA questionnaires was not directly accessible from the

StudentLife website, but was found within the dataset, in the file `EMA_definition.json` (Figure 9). As a JavaScript Object Notation (JSON) file, `EMA_definition.json` is a nested hierarchy of information that is both human-interpretable and computer-readable, which makes it particularly well-suited to programming and developing a data collection system like *StudentLife*. For example, when the *StudentLife* developers want to add a new question to the list of possible EMAs, they could simply write it into `EMA_definition.json`; in addition, when the student responses to the EMAs are collected in data, `EMA_definition.json` would help data analysts to understand what each of the responses mean. Most important to attend to in acknowledging this is how certain parts of the *StudentLife* data collection and processing system are more interpretable in this way, and are accordingly translatable or adaptable to other scenarios of use, such as data documentation in the case of `EMA_definition.json`. That EMA questionnaires leverage signifying semiotics more so than other data collection processes is an important factor to consider in why this occurs.

Altogether, the inclusion of `EMA_definition.json` in the *StudentLife* dataset might raise the question about why other ancillary files or scripts are excluded. This suggests that cleaning and deletion procedures beyond the scope of removing PII went into ensuring that the *StudentLife* dataset was publicly presentable. In this way, whereas *StudentLife* is a record of data, it is also a record of data collection and pre-processing procedures that ensured the dataset's presentability. Viewed in this way, the data analysis and application processes that *StudentLife* enables are constrained both by the data its data and its formatting, which is determined in part by data collection norms like removing PII.

Although the concern of this study is the *StudentLife* dataset, its operationalization by algorithms, and the rationalization of these algorithms, the *StudentLife* website that hosts and introduces this data cannot not be overlooked for rationalizing the collection, processing, and application of *StudentLife* data. For example,

```json
{
    "name": "Social",
    "questions": [
        {
            "options": "[1]0-4 persons, [2]5-9 persons, [3]10-19 persons, [4]20-49 persons, [5]50-99
            persons, [6]over 100 persons, ",
            "question_id": "number",
            "question_text": "How many people did you have contact with yesterday, including anyone you
            said hello to, chatted, talked or discussed matters with, whether you did it face-to-face, by
            telephone, by mail or on the internet, and whether you personally knew the person or not?
            Please select one of the following categories that best matches your estimate:"
        },
        {
            "options": "",
            "question_id": "location",
            "question_text": ""
        }
    ]
},
{
    "name": "Class",
    "questions": [
        {
            "options": "",
            "question_id": "course_id",
            "question_text": "What's the class name? (e.g., CS65)"
        },
        {
            "options": "[1]neutral, [2]strongly agree, [3]agree, [4]disagree, [5]strongly disagree, ",
            "question_id": "experience",
            "question_text": "I enjoyed the class today."
        },
        {
            "options": "[1]0, [2]1, [3]2, [4]3, [5]4, [6]5, [7]6, [8]7, [9]8, [10]9, [11]10, [12]>10, ",
            "question_id": "hours",
            "question_text": "How many hours did you spent on coursework outside class since the last
            class?"
        },
        {
            "options": "(Yes) 1 2 (No)",
            "question_id": "due",
            "question_text": "Do you have an assignment (due), quizz or exam today?"
        },
        {
            "options": "",
            "question_id": "location",
            "question_text": ""
        }
    ]
},
```

Figure 9: `EMA_definition.json` from the *Studentlife* dataset [102], with two multiple-choice EMA questions defined: `Social` and `Class`.

the first sub-heading on the *StudentLife* website frames the dataset according to a metaphor of "rhythm"—"Is there a rhythm to the Dartmouth term?" [100]—which, following Sandvig, we might read as metaphor that characterizes the *StudentLife* data collection system and dataset as objective or describable in physical terms. Rhythm in particular entails a natural order or cycle in Dartmouth student behaviors that rationalizes the motivation of the *StudentLife* project, which is articulated as an initiative to understand and reveal this order. Indeed, although finding a rhythm in Dartmouth student behavior according to the university's schedule is trivial, invoking a rhythm metaphor rationalizes *StudentLife* data collection, processing, and application by grounding these apparatuses in a claim that can be perceived as a matter of fact. On the *StudentLife* website, the rhythm metaphor is followed by a series of graphs that represent this rhythm graphically in a series of graphs (Figure 10). The lines of the curves make an argument that multiple data variables in the *StudentLife* dataset (e.g., gym, deadlines, mood) have rhythmic inter-relationships and might be accordingly interrelated or correlated.

Alongside statistical inferences derived from the *StudentLife* data is an embedded YouTube video entitled "Your Phone Knows Your GPA" [100], which describes the general motivations and outcomes of the *StudentLife* project. In the video, the *StudentLife* system is advertised as smartphone application that can sense student

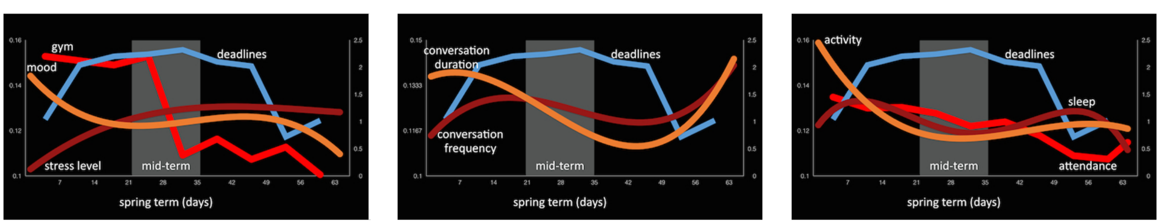

Figure 10: Visualizations of Dartmouth term "rhythms" on the *StudentLife* website [100], which depicts how certain *Studentlife* variables change over the course of the Dartmouth term in a correlative fashion.

behaviors and help them adjust these behaviors to improve their academic performance: “Imagine a world where a student simply checks their phone to see if their behavior is in sync with their desired GPA. If not, the phone provides tips on how to get back on track.” Because this directive is generally inconsistent with the goals of *StudentLife* described on the website and in publications, such as identifying “the impact of stress, mood, workload, sociability, sleep and mental health on academic performance” [100], it might be interpreted as one of many data applications of the *StudentLife* data. The “Your Phone Knows Your GPA” video altogether reflects a playful style and videography aesthetic that might not suit more formal academic outlets, and therefore should be interpreted as a different kind of rationalization than provided by the *StudentLife* academic publications. The video is perhaps less serious and intentional, but also might suit a different, ‘broader’ audience.

The video also summarizes results from the *StudentLife* study according to four statistical insights about “high performers”: they experience “an increase in stress levels up to the midterm period followed by a gradual decrease to the end of the term,” they “had shorter conversations during the evening and night periods later in the term,” they “spent more time studying,” and “they were more conscientious” (the latter point is accompanied by footage of a student making a to-do list). These statistical relationships between student behavior and academic performance warrant consideration as instantiating some premises of algorithmic governmentality. Whereas these inferred relationships between data variables are not applied directly to affect student behaviors, they are used to present an argument about the rationality of the *StudentLife* data collection system and dataset, which might be leveraged to correlate student behavior and academic performance.

The *StudentLife* website [100] includes “Publications” and “Presentations” sections, which list the academic publications and presentations in which the *StudentLife* dataset has been used or cited. This documentation is set alongside an “In Press” section that lists the news and academic organizations that have discussed *StudentLife*. Although it is without saying that these lists of citations and mentions rationalize the *StudentLife* project, what is more interesting to note is that these elements are privileged in the website structure over the data documentation, or a diagram of the dataset itself. The *StudentLife* dataset does not speak for itself, but is rather introduced according to a series of metaphors, explanations, and accreditations that rationalize its creation and application.

3.4.2 Design

In order to use or analyze the *StudentLife* dataset, I needed to further understand how each of its data variables could be related to one another and operationalized through algorithms and visualizations. This was the impetus for developing a data guide, which I referred to in order to experiment with simple statistical correlations between data variables.

Initial statistical experiments did not return anything interesting outside of what the *StudentLife* project had already identified: data variables that were correlated to high or low grades. It was also tedious to clean and pre-process each dataset in order for it to be used, which was the impetus for developing a more comprehensive data ontology. The ontology is structured according to the properties of the datasets, which supports their more efficient ‘ingestion’ by the *Lifestreams* data processing pipeline. At the highest level, the ontology is divided into temporal and non-temporal data, because whether data is recorded over time or whether it represents a subject’s unchanging trait, influences how it can be used. Whereas both temporal and non-temporal data are further divided into quantitative (e.g., numerical) and qualitative (e.g., text response) data, temporal data can also involve ‘event’ data, which is temporal but not necessarily quantitative nor qualitative. Location data in the *StudentLife* dataset is also temporal, and although it is either quantitative (e.g., GPS location) or nominal (e.g., wi-fi location), location data has special features that warrant its placement in a separate category. Within these sub-categories, each of the data variables from the data guide were manually sorted accordingly.

Whereas the data guide or documentation is a legible organization of the *StudentLife* data, the data ontology is a structured hierarchy that better suits data analysis and algorithmic operations. For example, data variables in the same ontological sub-categories can be handled in a similar programmatic manner, which expedites pre-processing. On the other hand, data variables in different categories need to be handled differently, which can reveal relationships between the data categories and how they might interact. For example, temporal qualitative data can be converted into temporal event data (e.g., a text response to a question is treated as an event), and temporal event data can be converted into quantitative data. (e.g., studying events can be converted to whether a student is studying or not at any given time). These inter-category relationships are crucial for relating data variables to one another or understanding the extent of their possible analysis. Much like a JSON file, a data ontology is like a compromise between a human-interpretable and a computer-readable structuration of data. Indeed, my pre-processing approach used the data ontology to output a `metadata.json` file, which *Lifestreams* would use to visualize individual data variables appropriately.

Data formats, values, and documentation all influenced the outcome of the data ontology; in turn, my plan for operationalizing this data was influenced by the ontology. Following from algorithmic governmentality, I noted how the non-temporal data category was conducive to developing subject 'profiles' [82]: the non-temporal data of grades and survey responses were trivially representable as profiles that could be compared to one another algorithmically. This insight made developing *Lifestreams* more tractable, because I could start from student profiles and 'build out' to other algorithmic inferences that incorporated more data types, like qualitative and temporal data. This pattern of data analysis was informed by the general constraints and capabilities of algorithms, which are most trivially applied to non-temporal quantitative data, and cannot so easily be applied to qualitative data (e.g., in *StudentLife*, survey and EMA questionnaire data), which is more heterogeneous. Although there are many ways to operationalize qualitative data, both the heterogeneity of the *StudentLife* data and the fundamental constraints of algorithms rendered this approach extraneous.

From the basis of student profiles derived from non-temporal data, and planning to 'build out' to other data types thereafter, I developed a data processing plan that would inform both the design and development of *Lifestreams*. The purpose of this plan was to successively incorporate data types into *Lifestreams*, but also to demonstrate these successive stages of data processing through the actual *Lifestreams* interface: for example, one part of *Lifestreams* could represent strictly non-temporal data in student profiles, and other parts could represent other data types. My initial plan was to separate each of these parts into distinct, successive phases of interaction, but this did not allow the different data types to be viewed alongside one another interactively. Accordingly, I designed each of the data types into separate components in the same visualization interface, which allowed their relationships to be revealed interactively. This design pattern is consistent with many visual analytic interfaces, which incorporate multiple visualizations into one interface instead of relying on one visualization technique to perform all analyses. It is a design paradigm that reflects the heterogeneity of data and the fact that data can be summarized and visualized in multiple ways.

Each step in designing how to implement data types into *Lifestreams* is elaborated as follows. Each step reflects categories in the data ontology, which served as a guide in their design. To be clear, these design considerations were determined before implementing the data into *Lifestreams*.

3.4.2.1 Non-Temporal Data

Student profiles can be created by incorporating each student's non-temporal data into a data object that represents them. If each of these data objects has the same structure, they can each be related to one another programmatically, which is the function of 'profiles' as identified by algorithmic governmentality. This in turn enables algorithmic inferences about a student's interests and predilections, which can be calculated with respect to the profiles of other students. Following from algorithmic governmentality and the directives of the *StudentLife* project, this would enable answering the questions: what students are related according to their data, how could these relationships be simplified to understand the student body better? Of particular interest here was GPA data, given the discursive presentation of the *StudentLife* dataset ("your phone knows your GPA"), but also questionnaire data that represented student opinions about political events on the Dartmouth campus. In particular, student opinions on campus protests and the Boston Bombing incident prompted by the *StudentLife* EMAs reflected an opportunity to understand how algorithms might operationalize sensitive opinion data with respect to other data variables.

Data profiles can be processed by a wide variety algorithms, but unsupervised machine learning algorithms support this study best because they can identify relationships between data profiles according to statistical relations between data variables. For example, 'clustering' algorithms can group student profiles into 'clusters' based on the similarity of their data. *Lifestreams*, however, uses 'dimensionality reduction' algorithms because they also calculate the relative importance of each data variable. This means that, whereas clustering algorithms treat all data variables as equally important, dimensionality reduction algorithms can operate to identify which data variables are statistically more useful than others. Furthermore, dimensionality reduction algorithms are particularly suited to expressing relationships between algorithms and visualization: because they can reduce a dataset of multiple data variables into a dataset of two data variables, the resulting two-dimensional data can be visualized on a two-dimensional plane. This allows for a human-interpretable depiction of algorithmic inferences, and importantly, without the overly suggestive output of 'clusters.' That is, for dimensionality reduction, groups among student profiles are not placed into discrete categories, but are rather made visible according to the relative proximities between data points. In this way, dimensionality reduction can be read as both a portrait of algorithmic governmentality—it programmatically instantiates a prioritization of statistical relationships between data variables—and also a caricature of mixed semiotics—it produces a 'de-territorialized' depiction of relations between data elements that evidences a-signifying semiotics.

The T-SNE [63] dimensionality reduction algorithm was selected in particular because it could operate in a high-dimensional space as demanded by the *StudentLife* dataset, because it can reflect multiple relationships in a high-dimensional space visually, and because its stochastic process reveals how the algorithm is working in real time. The PCA [96] dimensionality reduction algorithm is also included in *Lifestreams* for the sake of comparison with T-SNE, which enables comparing the algorithms' visual representations, their performance, and their applicability to certain data types. Fundamentally, the 'bias' of PCA is in favor of data variables that have the highest statistical 'variance' (the rationale being that data variables with a high variance characterize the data best), whereas T-SNE is suited to visualization because it privileges retaining relationships between each pair of data points, although irrespective of their overall arrangement.

In addition, an important design feature for *Lifestreams* was to give users the option to select data variables manually—without the assistance of machine learning and dimensionality reduction—in order to experiment with statistical inferences and to compare their selections with algorithmic ones. Such a focus was informed by FiveThirtyEight's "p-hacking" [5] widget (Figure 8), which allows its users to interact with variables in order to interactively design statistically significant correlations. But rather than demonstrating that dubious statistical inferences can be produced, *Lifestreams* intends to show how toggling data variables off and on determines the visual outputs of algorithms, which might be involved in rationalizing them. Altogether, dimensionality reduction, and especially T-SNE, is suited to such an experiment because it can yield varying and yet comparable visual results.

3.4.2.2 Temporal Data

Incorporating temporal data into the existing profiles of non-temporal data yielded a series of relationships between data variables that suggested some additional options for data analysis. Specifically, in addition to algorithmically inferring non-temporal predictions from non-temporal data by relating student profiles to one another, non-temporal data could be used to predict temporal data, and vice-versa. In addition, these inferential relationships could be manually adjusted by users as enabled by toggling data variables in dimensionality reduction. At this point I developed a metadata scheme that would programmatically indicate to *Lifestreams* where to include each of the data variables I decided to include. For example, GPA data would be nested under `features` (non-temporal data) as `[['gpa', ['total', 'semester', 'class']]`, whereas sleep data would be nested under `temporal` as `['sleep', 'semester', 0, 15]`, where `'semester'` indicates that the data spans an entire 10 weeks and `0` and `15` designate the minimum and maximum data values. Like `EMA_definition.json`, this `metadata.json` file operated as an interface between a human and programmatic interpretation of the data.

3.4.2.3 Location Data

Incorporating location data was more difficult than incorporating temporal data because, although *StudentLife* location data is formatted as temporal 'event' data, it is more informative as chronological 'sequence' data. This means that additional information can be derived from the order in which location events occur (sequences), or whether students have certain location chronologies in common (sequential patterns). To visualize these sequential patterns across all students, I chose to represent all of the location data as a directed graph or node-link diagram. A directed graph contains 'nodes' that are linked to one another by 'edges,' which can be used to represent locations (as nodes) and movements between them (as edges). Although a directed graph is not an algorithm per se, in *Lifestreams* it is derived algorithmically in two ways: first, data is algorithmically processed into a sequential format and then into a graph-based format; second, 'layout' algorithms determine how the nodes and edges should be visually arranged. Both of these algorithms can be modified, which influences not only how the directed graph is visualized, but also whether it reveals certain relationships between data variables or data profiles. Therefore, I planned for *Lifestreams* to allow users to customize one of these algorithms in a simple way.

3.4.2.4 Event Data

Event data, which is non-continuous temporal data, is relevant to a stage of 'data application' in algorithmic governmentality, because it entails the prediction and enactment of 'events' that affect subject behavior. That is, an 'event' can represent a behavioral circumstance that represents a subject's state (which might be predicted in advance), or can represent an action enacted in order to cause another event to (not) occur (as in 'just-in-time adaptive intervention' (JITAI) in healthcare [67]). Although these predictions and interventions could be simulated in *Lifestreams*, as discussed, data application departs from the reasonable scope of this study. Event-based predictions and interventions in particular involve a protracted relationship with the needs and concerns of data subjects that would be in poor taste to simulate or caricature in a speculative visualization. Nonetheless, more involved work in speculatively and critically examining the implications of event-based predictions and interventions should be pursued.

3.4.3 Development

With the data ontology and processing plan in place, I could begin developing the actual data pipeline that would ingest the *StudentLife* data into *Lifestreams*. I used a standalone Python script to pre-process the data variables and generate the metadata scheme, so whenever I wanted to incorporate new data variables or change their formats, I could edit and execute this script. I then used a JavaScript stack (Node[3], Webpack[4], React[5], D3[6]) to ingest the pre-processed data and render it as interactive visualizations. The development considerations and challenges for the specific visualization components are as follows.

3.4.3.1 Profiles

I began data analysis by aggregating non-temporal student data into profiles. For legibility, I represented each student profile as a data object that contained a nested representation of the data obtained from each student. Because some data was incomplete, certain students did not have certain data variables, which was reflected by missing elements in each student's data structure. However, in order to relate these profiles to one another algorithmically, they needed to be represented in an exactly consistent format. Accordingly, I added a function to the pre-processing Python script

[3]nodejs.org
[4]webpack.js.org
[5]facebook.github.io/react
[6]d3js.org

that would store all of each student's data variables in a single 'flattened' instance: an array of numbers, each of which related to a specific data variable. The function of what is called a 'data table' is the same as this: the indexical relationships between data variables (e.g., grades, survey responses) are preserved across data instances (i.e., students) in order to be more efficiently processed by algorithms.

Converting the data variables into a flattened, algorithmically-interpretable variable is a part of 'data cleaning.' If the flattened data was to remain strictly human-interpretable, it would be trivial to copy each data variable into its appropriate index. However, since the data variables needed to be computable, particularly according to dimensionality reduction, and more specifically by T-SNE, they needed to be specially formatted. Trivially, for example, whereas GPA can simply be stored in the flattened variable because it is already quantitative, nominal data obtained from multiple choice EMAs had to be quantified. In addition, class data was a particular challenge because it was stored as a list of classes for each student (e.g., Chemistry, Chinese, Psychology). In order to lend this data to dimensionality reduction, I used a 'one hot' approach to convert each class to a boolean variable for each student: e.g., if a student takes only Chemistry, then their Chemistry data variable is given a value of `1` (true), and their Chinese and Psychology data variables are both given a value of `0` (false).

Missing data also posed a particular challenge for simple statistical inferences in *Lifestreams*. For example, if missing GPA data was indicated numerically as `0` or `-1`, then it would affect the average GPA across all students. Accordingly, I had to programmatically ignore missing data (designated by `-1`) when calculating data averages across students. Moreover, data formatted using a 'one hot' approach was never treated as missing, because it reflects a boolean condition that is either true or false. For example, if a student had no class data, then they would be given a value of `0` (false) for every class that exists in the *StudentLife* dataset, whether or not it was true that the student actually took zero classes. This imposes an assumption about the data that is reflected in *Lifestreams*: the interface declares that the average value for having class COCS 065 is only 0.9, when in fact all students in the *Lifestreams* study should have been enrolled in COCS 065 [102]. This is because the COCS 065 is treated as either true or false, when in fact it is sometimes neither, in the event that data is missing.

3.4.3.2 Dimensionality Reduction

In order to produce interpretable visual patterns with T-SNE, I experimented with learning rates, iteration counts, and parameters that would determine how the data points would be arranged. To test the algorithm, I used synthetic data to ensure that similar data elements were positioned closer together. Since T-SNE is stochastic, groups of similar data are located in different places every time the algorithm is executed, but they should still retain their relative closeness or 'local' arrangements. The use of synthetic data to implement and debug T-SNE can be understood to demonstrate how algorithms can be agnostic to the signifying semiotics of data, which can refer to grades, stress, or random values and yet can still be operationalized. Similarly, like many other visual analytic systems, *Lifestreams* can be applied to other datasets so long as they are pre-processed into a similar format.

I implemented a selection feature that allowed me to highlight a group of data elements in the dimensionality reduction scatterplot. Incidentally, I could see how groups of data elements were preserved across executions of T-SNE (Figure 11). This feature was powerful not only for ensuring that T-SNE was running properly, but also for demonstrating how dimensionality reduction algorithms operate. In particular, by animating the algorithms' outputs between executions, and then by executing the algorithms any time a data variable was toggled off or on, animations between algorithm outputs revealed some of their underlying logics and biases (Figure 12).

Whereas each iteration of T-SNE's execution can be visualized in a way that animates the data elements, PCA does not involve iterations that lend themselves to animation, and so *Lifestreams* simply animates between the algorithm's outputs. Altogether, because PCA's output only changes when its input data variables are changed, and not every time the algorithm is executed, its animations are perhaps even better suited to revealing how the algorithm operates programmatically (compare Figures 11 and 12). Lastly, unlike T-SNE, PCA will not operate unless more than one data variable is selected, and so *Lifestreams* must prohibit this action with a pop-up alert.

3.4.3.3 Location Graph

To visualize the students' location data, I interpreted the `wifi_location` data as locational sequence data that could be rendered as a directed graph, where each wi-fi location corresponded to a node in the graph. This involved writing a simple algorithm that would iterate through the `wifi_location` data and store it in a graph-based data structure. Whereas each node in the graph represented a wi-fi location, each edge represented a student's movement from one location to another. The emergent effect of this approach is a map of all wi-fi locations in the data, which are

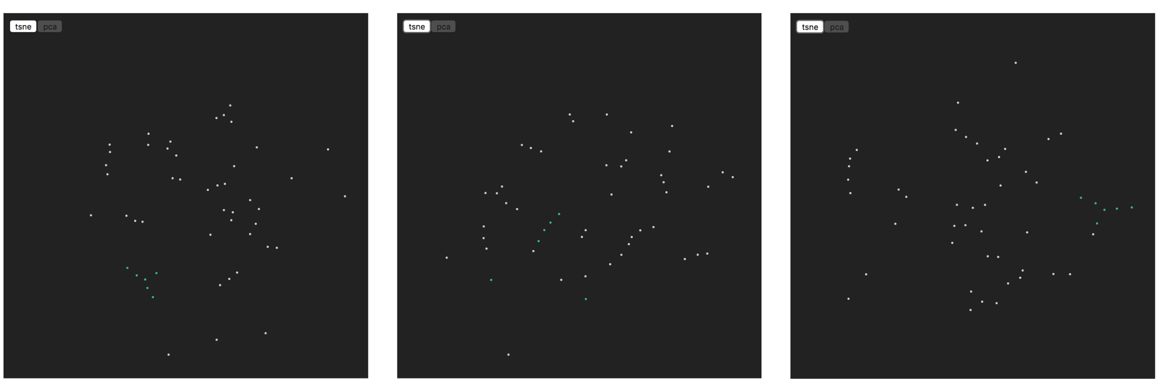

Figure 11: Every time the T-SNE dimensionality reduction is executed, it arranges the student data points differently. However, 'local' arrangements, like a selection of six students (left, highlighted green), are still preserved across iterations (middle and right).

connected according to student movements between them. Importantly, an inherent assumption here is that a student's wi-fi data is indicative of their location, which requires that all students possess their mobile devices at all times; where in fact, the *StudentLife* developers removed data at times during which the *StudentLife* app was deemed not on a student's person [102]. In addition, the *Lifestreams* algorithm that processes the wi-fi data naively assumes that any adjacent wi-fi locations represent a movement between these locations, which is untrue in the event that a student's mobile device is disabled or not connected to the Internet at a brief point in time. This is the reason that location data processing algorithm might be developed according to a spatiotemporal ontology, which would determine programmatically, to varying degrees of confidence, what patterns in data qualify as a student's presence at a location or movement between locations.

In comparison to GPS data, which consists of quantitative latitude and longitude positions, wi-fi data is nominal and so can be more easily grouped according to locational commonalities between students. Therefore, if *Lifestreams* used GPS instead of

Figure 12: When data variables are turned off in *Lifestreams*, the PCA scatterplot animates each time (left to right). Although some changes are minor (left to middle), other changes are more significant, like the horizontal axis being flipped (middle to right). This occurs because PCA recomputes which variables are statistically significant, updates the visualization accordingly, and does not make the axes reflect any data variables in particular. The axes are of a-signifying semiotics.

wi-fi location data to visualize student locations, additional data processing assumptions would be involved developing a spatiotemporal ontology. For example, one approach might be to divide geography into a grid, and determine student locations accordingly [105]. Altogether, for either wi-fi or GPS data, processing location data and defining spatiotemporal ontologies is always a matter of relating data to signifying semiotics in a process of literal territorialization. Although these territories can then be related to other data variables, operationalized, and thereby de-territorialized, contextualizing location data with signifying semiotics is generally essential.

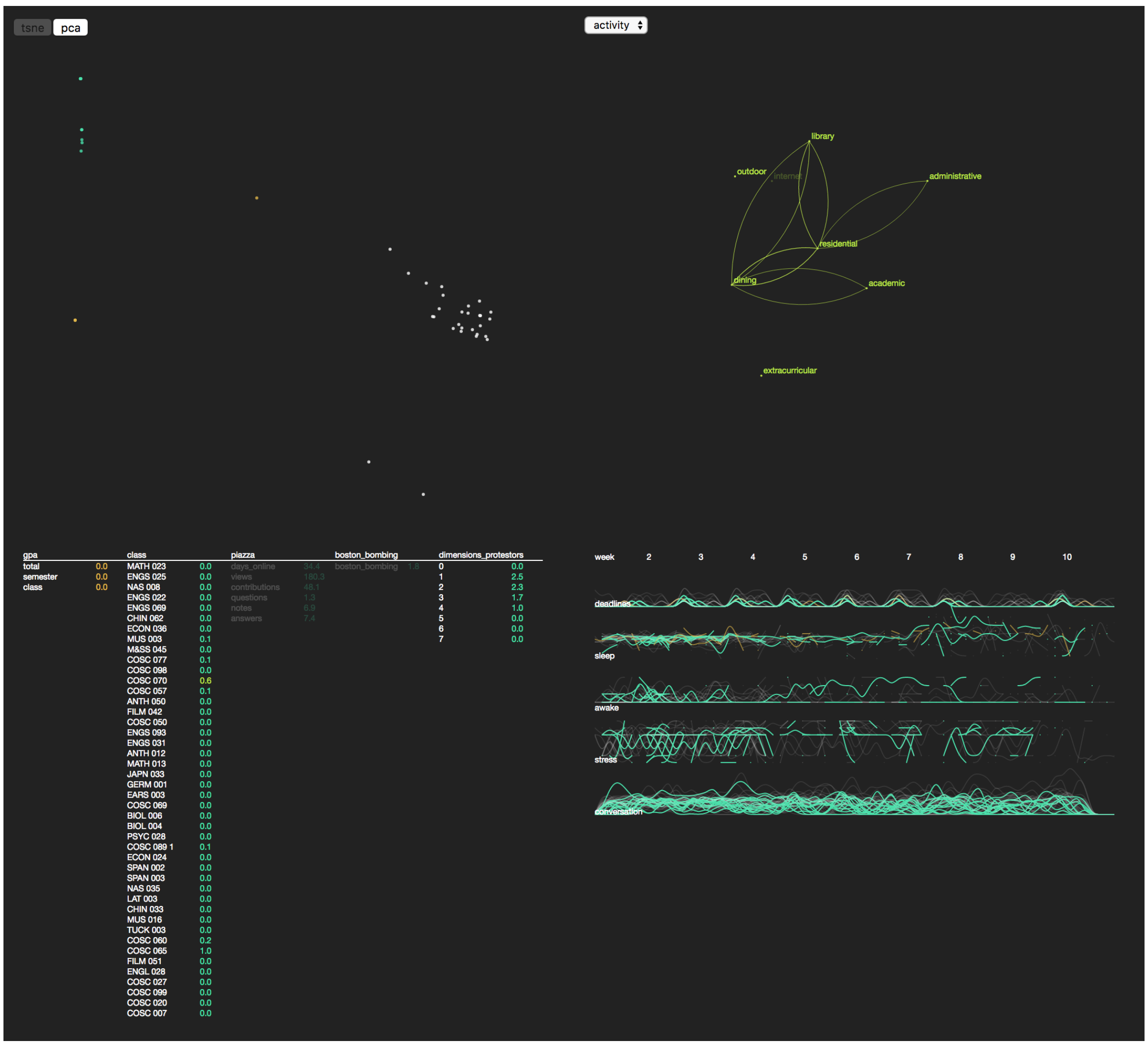

Figure 13: *Lifestreams* with the location graph (top right) set to the *activity* ontology, which collates network locations according to the behavioral activities that they entail.

To illustrate this, *Lifestreams* gives users three options to change the location graph according to very naive spatiotemporal ontologies: *network*, *college*, and *activity*. *Network* simply groups locations according to the names of wi-fi locations in the data, *college* groups locations according to which academic department they belong to, and *activity* (Figure 13) groups locations by the behavioral activity that they entail (e.g., academic, dining, extracurricular). In order to group by college and activity, I manually labeled every location in the *StudentLife* dataset according to information about each location that I gathered online. This was an insightful experiment in ascribing signifying meanings to data elements according to publicly accessible and sometimes subjective information. Whereas a veteran Dartmouth student might disagree with my classifications (e.g., a dining hall may actually be a popular studying location for students), an algorithm (or designer) that scrapes the internet to produce these labels would not incorporate these nuances.

Visualizing the directed graph of student locations required more work to ensure legibility than was required by the dimensionality reduction algorithms. Whereas the two-dimensional output of the dimensionality reductions can be plotted directly to a two-dimensional space, a directed graph has no necessary visual form and needs to be arranged according to a layout algorithm. Although I did not need to implement this layout algorithm from scratch (and used the D3 visualization programming language instead), I did need to manipulate it in order for the resulting visualization to be intelligible. In particular, I ensured that graph nodes retained a certain distance from one another, and that the nodes with the most frequented edges were closest together. In addition, I used curve edges to indicate bi-directionality, where moving clockwise along the curves indicates direction.

3.4.3.4 Timelines

Visualizing the temporal data as timelines was relatively simple. In order to compensate for missing temporal data, I used the D3 visualization programming language to `define` which parts of the timeline line should be displayed, and used a `round stroke-linecap` to reveal isolated temporal data points.

3.4.3.5 Interrelationships

I implemented 'brush-and-link' functionality between the visualization elements in order to reveal relationships between them. Trivially, for example, selecting students in the scatterplot highlights the data for those students in the location graph and the timeline (all active data in the data table is highlighted). But in order to reveal more informative, statistical relationships between data variables, I also included different layers of highlighting that indicated whether data variables were statistically related. In the current version of *Lifestreams*, when students are selected, other students that exhibit similar traits according to standard deviations are also highlighted. The location graph also uses a more sophisticated approach that highlights locations and movements that are especially characteristic of selected students: if a student is responsible for a certain percentage of traffic at a given location or route, then that location or route is highlighted. The intention in highlighting these statistical relationships is to reveal more relationships between data variables outside of what is already visualized in the dimensionality reduction scatterplot and via brush-and-link.

3.5 Discussion

Lifestreams presents a collection of interactive data visualizations developed in order to study and demonstrate the relationships between visualizations and rationalizations of algorithms, which is specifically informed by theories of algorithmic governmentality and mixed semiotics. To review, the algorithms implemented into

Lifestreams include ‘data-cleaning’ or pre-processing algorithms, dimensionality reduction algorithms (T-SNE and PCA), graph definition and layout algorithms (including *network*, *college*, and *activity* ontologies), and statistical correlations between data variables that are defined interactively. In addition, the *StudentLife* dataset ingested into *Lifestreams* entails an algorithmic data collection and processing apparatus that might also be rationalized by *Lifestreams*. To be clear, although every visualization and interaction library employed by *Lifestreams* involves algorithms, the ‘algorithms’ concerned here are those that have implications for algorithmic governmentality; that is, they collect and operationalize data about subjects.

After developing *Lifestreams* and collating notes about the development process, common themes were organized into a series of discussion points. Each discussion point was informed by considering *Lifestreams* in light of the categories of algorithmic governmentality—data collection, processing, application—as well as according to this study’s examples of mixed semiotic rationalizations of algorithms: gamification, visualization, recommendation, innovation, metaphorization, aestheticization. However, some of these considerations motivated multiple discussion points, some inspired none (considering the *Lifestreams* visualization in light of visualization is not very insightful), and almost all were collated, and so the three resulting topics no longer reflect the categories of algorithmic governmentality nor the examples of mixed semiotics. Instead, each is imagined as a genre of data analytics. For example, where visual analytics might concern the use of visualization to analyze data, democratic analytics might characterize the notion that democratic principles inhere in or are leveraged by data analysis. This rhetorical illustration has the effect of fragmenting the concept of data analytics and visual analytics into a series of sub-fields, each of which might be used as a lens to frame additional implications of data visualization and analysis.

3.5.1 Diegetic Analytics

Central to this study is the argument that algorithmic information systems like interactive visualizations present cohesive systems of meaning to users that shape their perceptions of the purposes, functions, and effects of algorithms. Both the 'participatory turn' and mixed semiotics support this claim, in that both help to identify instances of human-computer interaction in which the meanings of information systems seem inherent or immersive to users. Accordingly, *diegetic analytics* reads algorithmic information systems as immersive narratives that have diegetic and non-diegetic storytelling elements. Consider *Lifestreams*: the graphic and interactive elements of the visualization lend themselves to a directive of knowledge discovery or 'information foraging' [72], in which the user is actively engaged in finding information according to distributed visual and informational cues. Meanwhile, the possible consequences of finding this information, or operationalizing it in the future, are somewhat removed from the orienting graphics and interactions at hand—although the purpose of *Lifestreams* may be to promote the welfare of data subjects (following from the object of *StudentLife*), its immediate interactions cater more to the alchemical recombination of data variables and subjects, which is in order to ascertain some information about data. This is what the notion of machinic enslavement [59] attempts to address: that human-computer interaction can be automatic or reflexive in a way that makes it difficult for users to consider the implications of their interactions.

Therefore, when Rouvroy and Berns suggest that "inherent to contemporary statistical practice is the expression of individuals' tacit adhesion" [82], we might consider how presenting data subjects through *Lifestreams* expresses the *StudentLife* students as inherently conducive and consensual to the analytic operations being performed. That is, when *Lifestreams* is used, there is a tacit declaration that, firstly, the *StudentLife* data subjects are commensurate [30] to the analytic operations being performed, and secondly, that these subjects consent to the performance of these operations. Accordingly, it is unclear whether the diegetic framing of *Lifestreams* involves any consideration of human actors or individual subjects, because both the analyst and the subjects of *Lifestreams* are not represented in the diegetic space of visualization graphics and interactions. Instead, data points come to stand in for data subjects, correlations between variables take visual precedence over the meanings of these variables, and the *Lifestreams* user takes the role of a non-diegetic director, coordinating the movements of data subjects qua data points. In light of mixed semiotics, *Lifestreams* would thus seem to entail a departure from signifying semiotics that concedes to the de-territorialized, a-signifying semiotics of innumerable relationships between variables and subjects.

However, diegetic analytics comes with a caveat: if we compare interactive visualization to a diegetic frame, then we must also consider how this frame can rupture at any point, exposing its underlying, non-diegetic framing devices. For Amazon, the non-diegetic artifice of book recommendations is exposed whenever the recommendation algorithm malfunctions, but also when recommendations are inappropriate [3], or perhaps even when they are uncannily appropriate (e.g., how did Amazon predict that I wanted to purchase this book?). Accordingly, for *Lifestreams*, the inherent diegesis of knowledge discovery is compromised whenever discovered relationships between data subjects or variables seem farfetched, unrealistic, or even dire. Colloquial terms are useful for describing this paradigm: when something seems 'off' about the information suggested by *Lifestreams*, whether because it seems like a 'stretch' or because its implications are so novel that its sudden emergence is questionable, then the seamless coherence of *Lifestreams* is jeopardized. For this reason, diegetic analytics suggests that the salience of certain data analytic 'insights' exposes the non-diegetic artifice of data analysis, which opens it to scrutiny and critique, and which implies that machinic enslavement is not as totalizing as suggested by algorithmic governmentality.

The *Lifestreams* scatterplot evidences this claim, because although it presents salient relationships between data subjects according to a-signifying semiotics, the algorithmic calculation of these relationships is subject to scrutiny. That is, when *Lifestreams* positions data subjects in close proximity in the scatterplot, then salient relationships, patterns, and groups that emerge from this arrangement become meaningful a-signifying semiotics that can be referenced, discussed, or contested. The same scenario applies to the statistical similarities presented by highlighting data in *Lifestreams*: what is the underlying mechanism that suggests these data are related, and when these relations are insightful, how were they derived? This furthermore suggests that producing knowledge with algorithms entails a constant exchange between presenting a-signifying semiotics and rationalizing them according to other systems of meaning. Until a 'total' machinic enslavement or cybernetic 'singularity,' the a-signifying semiotics of algorithms are subject to critique without some dependency on signifying semiotics. In addition, the congruence or dissonance between signifying and a-signifying semiotics is always subject to contestation.

Accordingly, if Alexander Galloway's 'banality of representation' [36] suggests that the representational capabilities of visualizations are fundamentally limited, then we have implied here that this limitation is especially subject to critique. In *Lifestreams*, for example, that data subjects are represented as data points (per the banality of representation) is a configuration that might seem 'off' to users and thereby especially suspect. However, this study has also described how social norms—especially in the form of discursive and aesthetic regimes—contribute to whether the representations of visualizations seem uncanny. Therefore, certain users might perceive information systems (like *Lifestreams*) and their representations (like subjects-as-points) differently, some more enchanted or disenchanted by diegetic analytics. This altogether suggests that banal, routine, or a-political interactions with visualizations are subject to critique: they reveal a difference or discrepancy between the banality of representation and the completeness of its referents.

Effectively, diegetic analytics enables a departure from representationalist debates about whether visualization corresponds to reality 'or not' [66], and progresses instead to considering how these representations nonetheless produce knowledge about algorithms and their inferences. Following from Johanna Drucker [27], we might see this shift as acknowledging the "representational force" of visualizations, which although they are subjectively interpreted [28], produce knowledge. Incidentally, these considerations depart from the consternation of machinic enslavement and black boxes by acknowledging that algorithmic information systems, despite their effectiveness in producing knowledge and meanings, must also be rationalized according to meanings that suit individual subjects. A-signifying semiotics are manifest in data, algorithms, and visualizations, but to conceptually remove them from signifying semiotics is ultimately a stretch. Mixed semiotics is therefore particularly suited to identifying how these systems of meaning are exchanged or privileged in specific scenarios of algorithmic practice and use.

Indeed, using algorithmic information systems like *Lifestreams* allows for some negligence about the eventual implications of the knowledge discovered therein. And yet, even if the *Lifestreams* user is removed from the data application process, then the *StudentLife* developer or behavioral analyst that designs and tests the data applicator is less so. Similarly, if the *Lifestreams* user is removed from the meaning of the data, then the *StudentLife* data collector less so; if the *Lifestreams* user is removed from the logic of the machine learning algorithms, then whoever implemented these algorithms is less so. Indeed, who is most removed from this process is the *StudentLife* students, which would entail that the 'rarefication of subjectification' [82] in algorithmic governmentality is more of a 'displacement' or 'dispersion' of subjectification enabled by algorithmic paradigms of knowledge production. It is not so much that how algorithms produce knowledge is invisible or absolutely reflexive but that the considerations underlying these logics are distributed across subjects, mechanisms, and semiotics. Some considerations are delegated to the statistical reflexes of machine learning algorithms, but the question accordingly becomes when, where, why, and how: if algorithmic governmentality is not indefinite, everywhere, inevitable, or inherent, then we have additional opportunities for deciding how to interact with the algorithmic episteme.

3.5.2 Magnetic Analytics

When algorithmic governmentality suggests that its techniques "more effective because they are blurred," we might also witness how the roles of individual subjects are blurred by a 'displacement' or 'dispersion' of subjectification processes. Diverse constituents contribute to the algorithmic processes of *StudentLife* and *Lifestreams*, which provides these subjects each with a semblance of participation in algorithmic governance, and yet obfuscates exactly what the effect of their participation is. There may be signification and subjectification throughout these engagements—representations of the *StudentLife* mission in videos [100], compliance incentives for "top collectors" [102] —but the ambiguity of how these parts are integrated into *Lifestreams* caters to the illusion that all decisions originate instead from *Lifestreams*. Given that algorithmic information systems like *Lifestreams* are an outcome of de-territorialized data, logical, and semiotic elements produced by subjects and devices, they may operate to synthesize these parts into a cohesive "focal point" [59] of meaning. However, in acknowledging this, we must also recognize that this regime of meaning is not totalizing, totally enslaving, or totally indicative of 'the possible.' Accordingly, *magnetic analytics* imagines visualizations as attractive focal points that synthesize diverse bodies of information and action, and although they divert attention from the logic, materialities, and constituents of algorithms, to focus on them exclusively requires additional diversionary forces.

Interacting with *Lifestreams* in order to derive information from data prevents having to parse the *StudentLife* data manually, let alone in accordance with consulting the data documentation. Consequently, the value of *Lifestreams* is enhanced as a matter of economy: it becomes a one-stop shop for deriving information from the *StudentLife* data. Furthermore, when inferences are made in *Lifestreams*, it is absolutely unclear who or what is responsible for the specificity of the patterns revealed—e.g., why is this multiple-choice question on a scale of 7? why is standard deviation used here? or how does PCA work? In fact, revealing these considerations to users is the assignment of 'data provenance,' which is being increasingly incorporated into visual analytics systems: representing data provenance reveals both the history of data processing and its considerations, which can actually extend to data processing that is decided by users in real time [47]. And yet, data provenance is only a collection of data about data, which still supersedes consultation about why this meta-data was selected for in the first place, and by whom. For magnetic analytics, a function of what is called 'data provenance' is to avoid having to look 'beyond' the visualization display.

In magnetic analytics, all algorithmic inferences appear to resonate from *Lifestreams*, and yet nowhere do these algorithms have the first or the final say about how this data should be collected, processed, and applied. Indeed, Lifestreams draws connections between data subjects in a way that forgets the provenance of data collection and neglects the future of data application, but following from diegetic analytics, *Lifestreams* is not the full story—that it represents (amounts to) the entire *StudentLife* and *Lifestreams* assemblage is an optical illusion. A reductive interpretation of this might be to take the banality of representation, machinic enslavement, or black box discourse in order to argue that users are stuck in the meanings of algorithms or visualizations and absorbed into them. From this perspective, *Lifestreams* is the focal point that its users attend to in order to make decisions—it provides 'the possible.'

On the other hand, following from the participatory turn, mixed semiotics, and an aesthetics of algorithmic experience, *Lifestreams* is part and parcel of an enormous rationalization of an algorithmic apparatus, which also has recourse to videos and incentivization schemes—it is more an 'economy of the possible,' which incentivizes certain patterns of interaction over others, but it does not mandate them.

Therefore, magnetic analytics is enabled in part by the semiotic properties of data and algorithms (per algorithmic governmentality), but it is also rationalized by imbricating regimes of discourse and aesthetics, including those provided by visualizations. For example, Rouvroy and Berns cite how algorithmic "blindness to socially experienced categorizations (social, political, religious, ethnic, gendered, etc.) is the recurrent argument used by advocates of these algorithms replacing human evaluation" [106] (as cited in [82]). In other words, Lifestreams might be rationalized because it is not prone to the same human biases of psychologists or sociologists. However, following from the surveillance–innovation complex or an aesthetics of algorithmic experience, we would see how this "argument used by advocates of these algorithms" is one of discourse and aesthetics inasmuch as it leverages the semiotic affordances of algorithms. For example, the *StudentLife* study suggests that professors with "such data available" as *StudentLife* have a less reductive interpretation of student lives, and so prevents professors to judging students according to existing categorizations like GPA [102]. Therefore, although *StudentLife* and *Lifestreams* as algorithmic processes may be immune to certain categorizations, this feature in no way presents itself to users without some discursive or aesthetic rationalization of it, which is articulated in academic, journalistic, and public discourse.

Therefore, magnetic analytics pushes back against a conceptualization of algorithmic information systems as material agglomerations of data, mechanisms, and actors [58], opting instead for a sensitivity to the directionality or 'force vector' of arguments that rationalize them. Instead of conceiving *Lifestreams* as an object with an inherent meaning that is removed from the processes of its construction and hermeneutics, we interpret *Lifestreams* as an objectified deferral of human judgment and considerations—like a talking machine, a mechanical turk, or a Wizard of Oz. In this way, *Lifestreams* is not a peephole into algorithmic processes but a 'focal point' that is in open-ended dialogue with other forms of knowledge. It is the *attractive force* of this focal point that entices black box discourse, which for its part treats algorithms as all-or-nothing systems of meaning.

Following from mixed semiotics, then, we can begin to unravel how the system of meaning afforded by *Lifestreams* is *attractive* and not totalizing. *Lifestreams* is a system of mixed algorithmic, discursive, and graphic meanings that can be interpreted to understand *StudentLife* student behaviors. As such, when these semiotics cohere into salient patterns, *Lifestreams* operates as a *proof by assembly* that validates a particular construction of information. For instance, *Lifestreams* suggests that students with the highest overall grades are also those with less class assignment deadlines. That this relationship is revealed through *Lifestreams* suggests neither that this is inherently true nor that *Lifestreams* has all of the information necessary to make this claim (following from diegetic analytics), but that this data in composition with the *Lifestreams* system has a new meaning that, in its being revealed, enhances that value of all its parts.[7]

What is more is that, because *StudentLife* and *Lifestreams* both are capable of revealing potentially valuable insights such as these, either of their *proofs by assembly* suggest that *usable data is useful data.* This is not a tautology—it suggests that the collection, processing, and application of data rationalizes its utility. That a *collection* of 'sleep' data exists in *StudentLife* already rationalizes the value of this data,

[7] To be clear, although a *proof by assembly* might appear to originate from *StudentLife* or *Lifestreams*, following from mixed semiotics, it actually also a reflection of human valuations of coherence and efficiency, both of which inhere in visualizations. In other words, that the visualization 'speaks for itself' is an optical illusion, conceived here as an affordance of magnetic analytics.

which had to be designed, procured, and formatted. That this data is collated along with other data about 'grades' and 'stress,' accordingly, argues that these data variables can be meaningfully related. Therefore, this rationalization does not depend exclusively on informational insights that come from this data (like correlations), but also on the context provided by *StudentLife* or *Lifestreams* as a *proof by assembly*. That is, like appeals to 'open data,' *StudentLife* and *Lifestreams* appeal to 'assembled data,' which is a rationality with its own subjective, political, and semiotic parts that are not exclusive to the dataset or visualization itself. This is demonstrated, for example, by the *StudentLife* "opportunistic face logging" [101] research, which did not rationalize *StudentLife* through information contents but rather through information capabilities. More significantly, coupled with a 'displacement' or 'dispersion' of subjectification processes, *Lifestreams* confers to a structural harmony and unity of information that is efficient and economical to attend to when interpreting the *StudentLife* students. Supported by the participation of these students, the mission statement on the *StudentLife* website, and this thesis, *Lifestreams* presents a cohesive representation of so much labor, thought, and information that it is simply incredible.

3.5.3 Esoteric Analytics

Following from the banality of representation [36], *Lifestreams* is especially incredible because it incorporates more circumstances and information than can be explicitly represented in the visualization itself. The experiences and behaviors of *StudentLife* students, for example, are present in *Lifestreams* but not graphically represented in it, which is left to the user's imagination. In turn, following from an aesthetics of algorithmic experience [81] or metaphors of algorithms [86], *Lifestreams* compensates for this overabundance by representing information in these terms nonetheless, which confers to an illusion of completeness or comprehensiveness. That is, like on Amazon.com [57], signifying semiotics in *Lifestreams* represent (account for) algorithmic processes and a-signifying semiotics by which students and locations are graphically arranged. For *esoteric analytics*, this pattern of overabundance and reduction is leveraged to rationalize algorithms: as opposed to hiding or obfuscating the logics of algorithms in a black box, *Lifestreams* supplies them with signifying semiotics (students-as-points, Dartmouth building locations, data variables) as a frame of reference. Esoteric analytics is the underlying logic of the 'white box': *comprehensible feigns comprehensive.*

Esoteric analytics distracts from the implications 'little analytics' [2], which dismisses the totality of data in favor of statistical information and comprehension. Following from algorithmic governmentality [82], algorithms are semiotically effective at allowing this reductive motion to seem rational—their base traits and appearances allow for it. In turn, the *comprehensiveness* of these algorithmic insights is rationalized all the more when they are made explicitly *comprehensible*. For example, information derived from the *StudentLife* dataset is detailed and justified in academic publications. The function of *Lifestreams* in esoteric analytics is the same. Through diegetic analytics, *Lifestreams* presents an interactive 'information foraging' scenario in which all discovered relations seem to have a trail of interactive operations and graphic movements. Indeed, this trail of operations can be captured in 'insight provenance' [47], which would represent exactly how these discoveries were interactively achieved, thereby evidencing the knowledge discovery process. Then, through magnetic analytics, *Lifestreams* justifies its particular construction and rendition of evidence in a *proof by assembly*: that the *StudentLife* data variables and subjects can be related to one another in salient and coherent patterns rationalizes their mutual association. Together these effects are constitutive of esoteric analytics: they dispel an image of reductive 'little analytics' by showing comprehensibly—not comprehensively—how information was derived.

The question that is raised here is why esoteric analytics is permissible or preferable in the first place. It is not difficult to surmise a provisional answer. Firstly, that labor and resources are invested into designing and implementing *StudentLife* and *Lifestreams* demands their rationalization. For example, the *StudentLife* dataset is made publicly accessible online, which rationalizes the utility of the data in and of itself—following from either 'open data' [16] or 'assembled data' paradigms. Similarly, the time spent to develop *Lifestreams* is rationalized throughout this thesis. Secondly, and concurrently, these labor and resources implement infrastructures that engender standards in an "practical politics" [59]. For example, *Lifestreams* suits a particular kind of analysis that favors relationships between students and location data, and therefore might incentivize (and rationalize) location data collection. More significantly, that *StudentLife* publishes all of its data privileges any data analytic system that leverages the existing *StudentLife* data as opposed to collecting it from scratch, which would require more labor and resources.

The emergent effects of these 'practical politics' are what Ned Rossiter [79] terms a "determination of relevance," which characterizes the way in which existing standards (e.g., public *StudentLife* data) incentivize interoperable designs and actions (e.g., using *StudentLife* data in *Lifestreams*).[8] Following from an 'economy of the possible' [59], a determination of relevance is not strictly material, and it is simultaneously constitutive of and supported by mixed discursive and aesthetic semiotic regimes, which rationalize these standards in order to maintain their dependencies. For example, John Backus [6] notes how programmers in the 1950's obfuscated computer science as an arcane practice in order to build dependencies on their fluency,

[8]Such a pattern is evident also in disclosing patents for public domain, which effectively reinforces their value in a 'determination of relevance.' Alexander Galloway [38] and Tiziana Terranova [94] describe these dynamics according to internet infrastructure, and Michel Serres [88] generalizes them in his metaphysics.

which Wendy Chun [14] likens to a kind of esoteric source code "sourcery" to implicate contemporary algorithm scholars for doing the same. Such a paradigm is evident more broadly according to what R. Joshua Scannell [8] terms "digital mysticism," which is a prevailing rationalization qua obfuscation of algorithms that operates "on the grounds that the black box can, in fact, be deconstructed." Therefore, the function of esoteric analytics is to rationalize the standards and dependencies manifest in algorithmic information systems by representing their potential variety (e.g., of big data) as a unitary system of meaning (e.g., of little analytics, or of visualizations). To reiterate, magnetic analytics provides the centripetal force for this, and diegetic analytics retains it, both of which are a function of algorithm semiotics as well as discursive an aesthetic rationalizations.

What is additionally provided by esoteric analytics is a corrective regime in which system faults can always be attributed to other 'externalities' [79], thereby maintaining the sanctity of the system and its standards. For example, when *StudentLife* data is incorrect—perhaps a student forgot to report that they had homework deadlines one day—it the fault of the student as an externality. Similarly, "compliance" [102] is a subjectivating term that lends support to this hierarchy of accountability, in which a lack of adherence to following data collection protocols is attributed first to students and only incidentally to *StudentLife*'s design. Accordingly, when Rouvroy and Berns suggest that algorithmic governmentality "renders the very notion of misfire meaningless; in other words, a misfire cannot 'jeopardize' the system, it is immediately re-ingested to further refine behavioral models of profiles" [82], it is rather that the consequences of 'false positives' are exerted onto other constituents of the system that it does not depend on—they can either be swapped out or corrected by casting them as routine externalities. Esoteric analytics is precisely how this hierarchy of accountability is achieved: it represents dependencies in order to obfuscate the many ways that they might be otherwise conceived.

Altogether, esoteric analytics enables us to re-imagine black box *discourse* as discursive, aesthetic, and complicit in enabling a reductive understanding of the power and processuality of algorithms. The 'black boxed' algorithm is an objectification of structural power dynamics that exist both prior to and in support of algorithms—just like when algorithms are reductively caricatured as assembly lines or industrial machines [86]. The black box 'problem,' therefore, is not exclusively material but also discursive and aesthetic. Therefore, we err when we attempt to 'reverse engineer' [26, 18] algorithms, which are material manifestations of diverse design considerations, individual contributions, and subjective experiences. Such a reverse engineering approach is ultimately effective because it treats algorithms as scapegoats that provide us something discrete and finite to blame. And yet, the 'white box' it produces only obfuscates the actual construction and consequences of algorithms further. The black box and the visualization are both scapegoats in this way, and they are effective because they lend themselves to cohesive signifying semiotics.

Identifying an esoteric analytics suggests that there is, in fact, nothing magical about algorithms. More magical than algorithms are the narratives that inhere in their representations, aestheticizations, metaphorizations, and rationalizations. Such is to acknowledge that algorithms enable new paradigms of knowledge production, governance, and semiotics, and yet the difficulty in leveraging these capabilities as well as holding them accountable hinges on how algorithms are rationalized. By emphasizing the discursive and aesthetic regimes of algorithms, we depart from decrying the black box algorithm and bear witness to its non-diegetic construction and rationalization. Both visualization developers and critical scholars will play a part in this process, perhaps in working toward a non-diegetic analytics that brings our era of esoteric analytics to a close.

CHAPTER IV

CONCLUSION

This study implicates visualization in the rationalization of algorithms by investigating the design, development, and use of *Lifestreams*. It departs from a conceptualization of algorithms as entailing a pervasive, obtuse, black box hegemony in order to examine the importance of semiotic, discursive, and aesthetic regimes of meaning in rationalizing these algorithms. In this way, the study intends to broaden considerations of how visualizations represent algorithms. Algorithms are not just represented according to graphics (or enclosed within a black box); rather, interface narratives, metaphors, and conventions shape popular appraisals of algorithms in conversation with political discourse and legislation, metaphorical abstractions, and artworks. Therefore, attending to these rationalizations, studying them, and indeed designing them entails a politics of decision-making and knowledge production that remains under the jurisdiction of human actors.

Algorithms are increasingly inherent to contemporary society, and visual analytics is helpful for involving subjects in their operations. What we have seen is that subjects are included in and removed from certain points of the data collection, processing, and application pipeline in ways that are not so decisive as the black box metaphor would have it. And yet, power dynamics still inhere in these information systems whenever they produce knowledge. This does not entail the elimination of subjectification, nor total machinic enslavement, but it does blur the techniques of algorithmic governmentality, its subjects, and its interpretation. If anything, algorithms confer to a paradigm of human-computer interaction that rationalizes algorithms in different ways than techniques of ‘control and self-control’ are rationalized by governmentality proper. Algorithms do not just arrange data points or data subjects—they are enlisted to rationalize how these configurations are arranged.

REFERENCES

[1] Adorno, T. W., "On the fetish character in music and the regression of listening," *The essential Frankfurt school reader*, pp. 270–99, 1991.

[2] Amoore, L. and Piotukh, V., "Life beyond big data: Governing with little analytics," *Economy and Society*, vol. 44, no. 3, pp. 341–366, 2015.

[3] Ananny, M., "The curious connection between apps for gay men and sex offenders," *The Atlantic*, vol. 14, 2011.

[4] Arendt, H., *A Report on the Banality of Evil.* Faber, 1963.

[5] Aschwanden, C. and King, R., "Science isn't broken," *August*, vol. 11, p. 2015, 2015.

[6] Backus, J., "Programming in america in the 1950s–some personal impressions," *A History of Computing in the twentieth century*, pp. 125–135, 1980.

[7] Ballard, S., Mitew, T., Law, J., and Stirling, J., "Data natures: the politics and aesthetics of prediction," 2016.

[8] Behar, K., *Object-Oriented Feminism.* University of Minnesota Press, 2016.

[9] Blanchette, J.-F., "The noise in the archive: Oblivion in the age of total recall," in *Computers, Privacy and Data Protection: an Element of Choice*, pp. 25–38, Springer, 2011.

[10] Bogost, I., "Procedural rhetoric," *Persuasive games: The expressive power of videogames*, pp. 1–64, 2007.

[11] Bowker, G. C. and Star, S. L., "How things (actor-net) work: Classification, magic and the ubiquity of standards," *Philosophia*, vol. 25, no. 3-4, pp. 195–220, 1996.

[12] Chalmers, M. and MacColl, I., "Seamful and seamless design in ubiquitous computing," in *Workshop at the crossroads: The interaction of HCI and systems issues in UbiComp*, vol. 8, 2003.

[13] Chapman, O. and Sawchuk, K., "Research-creation: Intervention, analysis and" family resemblances"," *Canadian Journal of Communication*, vol. 37, no. 1, p. 5, 2012.

[14] Chun, W. H. K., "On "sourcery," or code as fetish," *Configurations*, vol. 16, no. 3, pp. 299–324, 2008.

[15] Clough, P. T., Gregory, K., Haber, B., and Scannell, R. J., "The datalogical turn," *Non-Representational Methodologies: Re-Envisioning Research*, vol. 12, p. 146, 2015.

[16] Cohen, J. E., "The surveillance-innovation complex: The irony of the participatory turn," 2014.

[17] Cuevas, H., Hatzopoulos, P., Huws, U., Kambouri, E., Mezzadra, S., Neilson, B., Rossiter, N., Samaddar, R., and Walters, W., "logisticalworlds," 2014.

[18] Datta, A., Sen, S., and Zick, Y., "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *Security and Privacy (SP), 2016 IEEE Symposium on*, pp. 598–617, IEEE, 2016.

[19] Davidson, D. and Mateas, M., "Procedural literacy: Educating the new media practitioner," *On the Horizon*, vol. 13, no. 2, pp. 101–111, 2005.

[20] Davila, P., *Visualization as Assemblage: How Modesty, Ethics, and Attachment Inform a Critical Design Practice*. PhD thesis, 2016.

[21] de Zeeuw, Daniel, Z. S. and Rossiter, N., "Of pirarchy, anonimity, and parametric politics: an interview with ned rossiter and soenke zehle," 2015.

[22] Deleuze, G., "Postscript on the societies of control," *October*, vol. 59, pp. 3–7, 1992.

[23] Deleuze, G., Guattari, F., and others, *A thousand plateaus*. Minneapolis: University of Minnesota Press, 1987.

[24] Design, S. N., "Nike clc installation."

[25] Designboom, "Super nature design: Nike china logistics center," 2012.

[26] Diakopoulos, N., "Algorithmic accountability: Journalistic investigation of computational power structures," *Digital Journalism*, vol. 3, no. 3, pp. 398–415, 2015.

[27] Drucker, J., "Humanities approaches to graphical display," *Digital Humanities Quarterly*, vol. 5, no. 1, pp. 1–21, 2011.

[28] Drucker, J., "Performative materiality and theoretical approaches to interface," *Digital Humanities Quarterly*, vol. 7, no. 1, 2013.

[29] Drucker, J. and Nowviskie, B., "Speculative computing: Aesthetic provocations in humanities computing," *A companion to digital humanities*, pp. 431–447, 2004.

[30] Espeland, W. N. and Stevens, M. L., "Commensuration as a social process," *Annual review of sociology*, vol. 24, no. 1, pp. 313–343, 1998.

[31] Evans, L., "How to build a map for nothing: Immaterial labour and location-based social networking," 2013.

[32] Foucault, M., *Madness and civilization: A history of insanity in the age of reason.* Vintage, 1988.

[33] Foucault, M., *Security, territory, population: lectures at the Collège de France, 1977-78.* Springer, 2007.

[34] Foucault, M., *Discipline & punish: The birth of the prison.* Vintage, 2012.

[35] Fuller, M. and Matos, S., "Feral computing: From ubiquitous calculation to wild interactions," *The Fibreculture Journal*, no. 19 2011: Ubiquity, 2011.

[36] Galloway, A., "Are some things unrepresentable?," *Theory, Culture & Society*, vol. 28, no. 7-8, pp. 85–102, 2011.

[37] Galloway, A., "The poverty of philosophy: Realism and post-fordism," *Critical Inquiry*, vol. 39, no. 2, pp. 347–366, 2013.

[38] Galloway, A. R., *Protocol*, vol. 23. SAGE Publications London, Thousand Oaks, CA and New Delhi, 2006.

[39] Gaylor, B., "Do not track," 2015.

[40] Genosko, G., "Félix guattari," *Angelaki: Journal of Theoretical Humanities*, vol. 8, no. 1, pp. 129–140, 2003.

[41] Genosko, G., "Information and asignification," *FOOTPRINT*, vol. 8, no. 1, pp. 13–28, 2014.

[42] Genosko, G. and Bouissac, P., *Critical Semiotics: Theory, from Information to Affect.* Bloomsbury Publishing, 2016.

[43] Gillespie, T., Gillespie, T., Boczkowski, P., and Foot, K., "The relevance of algorithms," *Media technologies: Essays on communication, materiality, and society*, vol. 167, 2014.

[44] Gillespie, T. and Seaver, N., "Critical algorithm studies: A reading list," *Socialmediacollective. org*, 2016.

[45] Google, "How search works: From algorithms to answers," 2013.

[46] Google, "Google dictionary," 2017.

[47] Gotz, D. and Zhou, M. X., "Characterizing users' visual analytic activity for insight provenance," *Information Visualization*, vol. 8, no. 1, pp. 42–55, 2009.

[48] Green, D. P., Ha, S. E., and Bullock, J. G., "Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose," *The Annals of the American Academy of Political and Social Science*, vol. 628, no. 1, pp. 200–208, 2010.

[49] Guattari, F., *Molecular revolution: Psychiatry and politics.* Penguin Group USA, 1984.

[50] Haim, S., Loeb, L., Wang, R., Zhou, X., Lord, S., and Campbell, A., "The mobile stress meter: A new way to measure stress using images," ACM, 2015.

[51] Hajian, S., Bonchi, F., and Castillo, C., "Algorithmic bias: From discrimination discovery to fairness-aware data mining," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2125–2126, ACM, 2016.

[52] Jenkins, T. and Bogost, I., "Designing for the internet of things: prototyping material interactions," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 731–740, ACM, 2014.

[53] Khoury, M. J., McBride, C. M., Schully, S. D., Ioannidis, J. P., Feero, W. G., Janssens, A. C. J., Gwinn, M., Simons-Morton, D. G., Bernhardt, J. M., Cargill, M., and others, "The scientific foundation for personal genomics: recommendations from a national institutes of health–centers for disease control and prevention multidisciplinary workshop," *Genetics in Medicine*, vol. 11, no. 8, pp. 559–567, 2009.

[54] Kittler, F., "Protected mode," *Draculas Vermachtnis: Technische Schriften*, pp. 208–24, 1993.

[55] Kowark, T., Muller, J., Muller, S., and Zeier, A., "An educational testbed for the computational analysis of collaboration in early stages of software development processes," in *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pp. 1–10, IEEE, 2011.

[56] Krishnamurthy, B. and Wills, C. E., "On the leakage of personally identifiable information via online social networks," in *Proceedings of the 2nd ACM workshop on Online social networks*, pp. 7–12, ACM, 2009.

[57] Langlois, G., *The TechnoCultural Dimensions of Meaning.* PhD thesis, Citeseer, 2008.

[58] Latour, B., "From aggregation to navigation: A few challenges for social theory," *Closing Keynote Plenary, CHI*, vol. 13, 2013.

[59] Lazzarato, M. and Jordan, J. D., *Signs and machines: Capitalism and the production of subjectivity.* Semiotext (e) Los Angeles, 2014.

[60] Lingenfelter, D. and Khurshudov, A., "The black box paradox in big data analytics and data-driven modeling," 2014.

[61] Loukissas, Y. A., "Taking big data apart: local readings of composite media collections," *Information, Communication & Society*, vol. 20, no. 5, pp. 651–664, 2017.

[62] Lukens, J. and DiSalvo, C., "Speculative design and technological fluency," *International Journal of Learning and Media*, 2012.

[63] Maaten, L. v. d. and Hinton, G., "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[64] Mordvintsev, Alexander, O. C. and Tyka, M., "Deepdream: a code example for visualizing neutral networks," 2015.

[65] Murphy, D., "Nict's daedalus creates beautiful 3d visuals to map out nasty cyber attacks (video)," *Engadget*, 2012.

[66] Nacher Dr, A., "Internet of things and automation of imaging: beyond representationalism," *communication+ 1*, vol. 5, no. 1, p. 6, 2016.

[67] Nahum-Shani, I., Smith, S. N., Tewari, A., Witkiewitz, K., Collins, L. M., Spring, B., and Murphy, S., "Just in time adaptive interventions (jitais): An organizing framework for ongoing health behavior support," *Methodology Center technical report*, no. 14-126, 2014.

[68] Norman, D. A., *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[69] Parikka, J., *What is media archaeology?* John Wiley & Sons, 2013.

[70] Pasquale, F., *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015.

[71] Pentland, A., *Social physics: How good ideas spread-the lessons from a new science*. Penguin, 2014.

[72] Pirolli, P. and Card, S., "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[73] Projects, Y. I., "Nike+ collab: City runs."

[74] Puschmann, C. and Burgess, J., "Big data, big questions| metaphors of big data," *International Journal of Communication*, vol. 8, p. 20, 2014.

[75] Rancière, J., *Aisthesis: Scenes from the aesthetic regime of art*. Verso, 2013.

[76] Ratto, M., "Critical making: Conceptual and material studies in technology and social life," *The Information Society*, vol. 27, no. 4, pp. 252–260, 2011.

[77] Reidsma, M., "Ethical ux," 2017.

[78] Reigeluth, T., "Why data is not enough: Digital traces as control of self and self-control," *Surveillance & Society*, vol. 12, no. 2, p. 243, 2014.

[79] Rossiter, N., *Software, Infrastructure, Labor: A Media Theory of Logistical Nightmares*. Routledge, 2016.

[80] Rossiter, N. and Zehle, S., "Locative media as logistical media: Situating infrastructure and the governance of labor in supply-chain capitalism," 2015.

[81] Rossiter, N. and Zehle, S., "The aesthetics of algorithmic experience," *The Routledge Companion to Art and Politics*, pp. 214–221, 2015.

[82] Rouvroy, A., Berns, T., and Libbrecht, E., "Algorithmic governmentality and prospects of emancipation," *Réseaux*, no. 1, pp. 163–196, 2013.

[83] Rouvroy, A., Hildebrandt, M., and De Vries, K., "The end (s) of critique," *Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology*, p. 143, 2013.

[84] Rouvroy, A. and Stiegler, B., "The digital regime of truth: From the algorithmic governmentality to a new rule of law,"

[85] RYBN, "Antidatamining," 2017.

[86] Sandvig, C., "Seeing the sort: The aesthetic and industrial defense of "the algorithm"," *Journal of the New Media Caucus*, 2014.

[87] Seitz, A., "It's a series of tubes: Network neutrality in the united states and how the current economic environment presents a unique opportunity to invest in the future of the internet," *J. Nat'l Ass'n Admin. L. Judiciary*, vol. 29, p. 683, 2009.

[88] Serres, M., *Le parasite*. Grasset, 1980.

[89] Smilkov, D., "A neural network playground," 2017.

[90] Star, S. L. and Ruhleder, K., "Steps toward an ecology of infrastructure: Design and access for large information spaces," *Information systems research*, vol. 7, no. 1, pp. 111–134, 1996.

[91] Stone, A. A. and Shiffman, S., "Ecological momentary assessment (ema) in behavorial medicine.," *Annals of Behavioral Medicine*, 1994.

[92] Striphas, T., "Algorithmic culture," *European Journal of Cultural Studies*, vol. 18, no. 4-5, pp. 395–412, 2015.

[93] Swan, M., "The quantified self: Fundamental disruption in big data science and biological discovery," *Big Data*, vol. 1, no. 2, pp. 85–99, 2013.

[94] Terranova, T., *Network culture: Politics for the information age.* Pluto Press, 2004.

[95] Thomas, J. J., *Illuminating the path:[the research and development agenda for visual analytics].* IEEE Computer Society, 2005.

[96] Tipping, M. E. and Bishop, C. M., "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[97] Tiqqun, *The Cybernetic Hypothesis.* 2010.

[98] Tsavli, M., Efraimidis, P. S., Katos, V., and Mitrou, L., "Reengineering the user: privacy concerns about personal data on smartphones," *Information & Computer Security*, vol. 23, no. 4, pp. 394–405, 2015.

[99] Victor, B., "Explorable explanations," *Bret Victor*, vol. 10, 2011.

[100] Wang, R., "Studentlife study," 2017.

[101] Wang, R., Campbell, A. T., and Zhou, X., "Using opportunistic face logging from smartphone to infer mental health: challenges and future directions," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 683–692, ACM, 2015.

[102] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T., "Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 3–14, ACM, 2014.

[103] Wang, R., Harari, G., Hao, P., Zhou, X., and Campbell, A. T., "Smartgpa: how smartphones can assess and predict academic performance of college students," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pp. 295–306, ACM, 2015.

[104] Widman, J., "Presenting edgerank: A guide to facebook's newsfeed algorithm," 2011.

[105] Yuan, M. and Nara, A., "Space-time analytics of tracks for the understanding of patterns of life," in *Space-Time Integration in Geography and GIScience*, pp. 373–398, Springer, 2015.

[106] Zarsky, T. Z., "Governmental data mining and its alternatives," *Penn St. L. Rev.*, vol. 116, p. 285, 2011.

[107] Zehle, S. and Rossiter, N., "Algorithmic architectures, logistical media, and the rise of black box politics," *The Routledge Companion to Labor and Media*, p. 40, 2015.

[108] Zimmerman, J., Forlizzi, J., and Evenson, S., "Research through design as a method for interaction design research in hci," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 493–502, ACM, 2007.